MACMILLAN AND CO., LIMITED
LONDON · BOMBAY · CALCUTTA · MADRAS
MELBOURNE

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO
DALLAS · ATLANTA · SAN FRANCISCO

THE MACMILLAN COMPANY
OF CANADA, LIMITED
TORONTO

The Technique
of
Examining Children

A Quest for Capacity

By
B. C. Wallis

MACMILLAN AND CO., LIMITED
ST. MARTIN'S STREET, LONDON
1930

First Edition 1927.
Reprinted 1930.

PRINTED IN GREAT BRITAIN

PREFACE

MUCH misunderstanding prevails concerning scholarship examinations. This little book aims at clearing away some common misconceptions. It is based upon an experience extending continuously over more than twenty years as an examiner, and during recent years as a chief examiner for scholarships in three important areas—London, Birmingham, and Brighton. It is not an account of the system which obtains in any one examination scheme, but represents a generalised statement of the methods which this lengthy experience has suggested and of the principles which these methods embody.

It is confidently anticipated that this book will give rise to criticisms from many points of view; otherwise it were not worth the labour of writing. 'Chapter and verse' are not given for certain facts either specifically mentioned or implied; the initiated will recognise the sources, and to others the information is of little value.

One acknowledgment must be made. It is not possible to estimate how much I owe to the fact that my predecessors in London had developed a system which was working quite smoothly and fulfilling its main purpose to the general satisfaction of all those whom it concerned.

As I hold the position of Chief Examiner for Scholarships for the Education Authorities of London, Birmingham and Brighton, it is necessary to state prominently and explicitly that this book carries no official sanction. None of these authorities is, in the smallest degree, responsible for any statement in the book. I have been able, fortunately, to quote from papers set in connexion with the examinations conducted by these authorities ; the papers in the Appendix were all set in one of these three places.

My grateful thanks are due to my London colleagues, Mr. A. J. Williams and Mr. H. Grigs, for useful suggestions which arose from their kind reading of the proofs.

B. C. WALLIS.

LONDON, *January*, 1927.

CONTENTS

I. TRADITIONS IN EXAMINING

Introductory. Some form of examination or test is inevitable in the educational life of England. Based upon the spirit of competition and dominated by the sense of fair-play, some method of selection is fundamental to our national life. On the average and in the long run the Heads of schools can label with fair precision the quality of their scholars, but he would be rash who would assert that the best pupil in 'Trafalgar' school was inevitably and always a better pupil than the best pupil in 'Waterloo' school. While Heads may have a general notion of the capacity of their children in relation to the national average, they are not in the least degree able to arrange in order of merit their own children in open competition with the children of, say, half a dozen neighbouring schools. Hence the need for an external test and a strictly fair competitive examination.

The employer who buys the services of youths and maidens in the open market has the right to demand that the capacities of these prospective employees shall be gauged with relative precision on a common standard. The examination test must, therefore, be competitive either by pupil against pupil or by pupil

against a common norm, the hypothetical average pupil of normal capacity.

Despite many drawbacks, none of which is essential, examining is at worst a necessary evil, and at best a profound stimulus ; hence this brief outline of one conception of the form and methods of modern examining ; and hence, also, as a beginning, a brief discussion of examining traditions, most of which are worn out and, in a modern view, either unnecessary or harmful. Six of the main traditions are treated in the paragraphs which follow.

I. Secrecy. It is a tradition that examination methods, principles, and machinery should be surrounded with mystery. Secrecy is the child of fear ; and, probably, no examiner has completed his estimate of the ability of a batch of candidates without a decided dread that he has been unfair to one or other of them. Probably the greater the experience the more insistent the dread ; certainly none but the novice feels a supreme confidence in the accuracy of his assessment. The over-conscientious examiner is, at times, oppressed by this ever-present dread of unfairness.

Out of any batch of candidates some are almost sure to give vent to their dissatisfaction at the examination result, and examining authorities erect a barrier of secrecy against such disgruntled people who are unable to ' play the game.'

But when the benefits have been duly appreciated the contra account is sufficiently weighty to be an argument for the greatest possible publicity. Without entering into details it is certain that in the past

the cloak of secrecy has at infrequent intervals covered deliberate and culpable neglect of duty, gross inefficiency, as well as other defects due to the bias or caprice or deliberate intention of examiners.

The movement for reforms in examining which has been in progress for the last decade would have yielded its results more rapidly and more easily if the technique and principles of examining had not been deliberately obscured.

Further, assistant examiners have had a legitimate excuse for a lack of interest in their work, as information concerning the character of their assessments has been, and is, withheld. Assistant examiners habitually 'carry on' under considerable disabilities, not the least of which is ignorance of how their work turns out. Most craftsmen have some knowledge of the finished product to which they have contributed; the assistant examiner, working as one of a team, marks scripts labelled by numbers, and is not informed of the final issue of his labours or of the relationship which his efforts bear to those of his co-workers.

II. Infallibility. Happily the tradition of the infallibility of the examiner was questioned some years ago. Formerly, as part of the general secrecy, there was no appeal from the examiner's verdict, yet experiment after experiment demonstrated the facility with which examiners of repute varied in their judgments. Although it is difficult to devise a workable court of appeal, and although authorities must decline on these grounds to revise the published examination results, the whole trend of modern

examining technique is to ensure the candidate against failure by virtue of the personality of the examiners. In practice examiners tend to a mechanical accuracy within a small margin of error. This in itself is a notable consequence of the enquiry into examination methods to which reference has already been made. While individual examiners, being human, are, as a class, neither more nor less fallible than their predecessors, the net result of the examining system is sufficiently precise, within the limits of error suitable to a particular examination, for the candidates to feel confidence in the result. Failure or success has not been *awarded* by the examiners as a consequence of the reaction between the candidate and the examiners, but has been *scored* by a precise and dispassionate assessment of the work which the candidate has done in the examination room. The examiner merely records the candidate's score on the target.

III. Every teacher a good examiner. Teachers are continually weighing and assessing the performances of children ; their experience in estimating the value of the written word, in scoring a mark for an exercise in English or Arithmetic, is necessarily extensive. It has become a tradition that an examiner should be, or have been, a teacher ; in fact, it has been argued with some cogency that the only examiner to whom the work of a child should be submitted is the child's teacher. Critical experience demonstrates that the teacher is constrained by his teaching experience to be a bad examiner. In the first place the good teacher has a personality which gives a distinct bias

to his teaching, and he habitually assesses as good the *style* of performance which he prefers. Undoubtedly his bias affects him unconsciously, and an examiner with a bias is a bad examiner, the more so if he be unconscious of the way in which the bias trends.

Secondly, in marking class exercises merit is not the only consideration which determines the score. Some teachers fail to differentiate critically between the good and the very good, and have a favourite mark—9 or even 10 (out of 10)—which they *bestow* on work which they regard as satisfactory; others lump their awards together—5, 6, and 7 (out of 10)—in some 90 per cent. of the cases. Most teachers develop idiosyncrasies in their class marks, and few teachers are in the habit of *awarding* marks at all points on the scale. These habits depend, in some cases, upon a tradition that it is bad for a child to receive *full* marks, and, in other cases, upon a desire to stimulate or restrain the activities of a particular pupil. Class marks are an extraordinarily useful aid to teaching when they are skilfully used, but the successful practitioner becomes by reason of his success a bad examiner. It is a matter of considerable difficulty to persuade examiners who are teachers to score 75 per cent. or over, and 25 per cent. or less in an adequate number of cases. Analysis has shown that an examiner who had gained a considerable reputation owed his apparent success as a consistent examiner to his inability to award marks other than 3, 4, 5, 6, 7, out of 10. Finally, a teacher tends to lose plasticity of mind

and, as an examiner, is prone to penalise candidates for habits which he considers are faults, but which are due entirely to the teaching to which the candidates have been submitted.

An examiner meets the following circumstances not infrequently: the scripts of a batch of candidates from the same form in a school give definite evidence throughout of rank bad teaching, *i.e.* the teaching of errors, or of misguided and misdirected teaching. The candidates are obviously capable; the fault lies almost solely with the teacher, who may be, for example, an indifferent disciplinarian. Freedom from prejudice and an habitually open and receptive mind are fundamental to an examiner in such a case. An examiner is not primarily a critic; he is a human machine which receives and records impressions from candidates' scripts readily and consistently. He is a searcher, an investigator with an entirely open mind. A teacher examiner must forget the classroom and the methods he uses and prescribes therein.

IV. One purpose of examining is to improve teaching. This tradition will yet live long. Examiners' reports usually refer to teaching; frequently the examiner is directed to report on the teaching and not on the performances of the candidates. Teachers of repute, specialists in a given department, have not infrequently sought for, and received, appointments as examiners in order that they may influence teaching along their own special lines. An assistant examiner who has pleaded for

freedom and breadth of outlook has been liable to be checked by an answer to his plea that a narrow interpretation of his instructions is alone permissible. A chief examiner with a definite purpose is a bad examiner. It has, even, happened that a nominally general examination has become in practice a test to pick out a few candidates with a special aptitude.

Probably the best evidence that this tradition still flourishes is to be found in the remarkably variable percentage of passes in different subjects at different examinations. Teachers of experience know that such and such a subject at one examination is a 'soft option,' and at another examination is a formidable barrier. Time and again teachers submit candidates in those subjects which their experience indicates are easy. The examiner in the difficult subjects is trying to bias the teaching, and the unfortunate candidate suffers because his teacher and his examiner are at loggerheads.

There is no excuse whatever for using a *general* examination as an indirect method of influencing the teaching in the schools. A *general* examination should follow the teaching, and neither guide it nor dictate to it. This dictum is based entirely and frankly upon the notion that a *general* examination is intended to grade candidates in accord with their inherent capacity. A *general* examination condemns itself when it reports that the teaching has improved: the capacity of the candidates, provided they are sufficiently numerous year after year, has not changed, and either in the past or in the present

(possibly all the time) the examination has been a failure.

V. An examination is based upon what a candidate ought to know. The extreme cases due to the tradition that a general examination has a pre-judged standard of attainment or knowledge have been already referred to, but the general application of the tradition requires some consideration. In practice the tradition works badly. For the general schools examination, for example, the assistant examiner is expected to be able readily to recognise a 'pass' answer, *i.e.* an answer which should score half marks; in fact, it would appear that the chief qualification of an assistant examiner is precisely this ability. The assistant examiner is expected when he first reads an examination question to be able to settle *à priori* the answer which shall pass. Probably the assistant examiner has little difficulty in making up his mind as to what constitutes, in his opinion, a 'pass' answer, but he has no guarantee that his colleagues would concur with his verdict. In competitive scholarship examinations the tradition works badly in this wise. With his prejudice towards what a candidate ought to do and ought to know, the assistant examiner finds great difficulty in awarding good marks for answers which do not reach his prejudiced standard. For example, when a question proves difficult for a majority of the candidates, the assistant examiner finds an equal difficulty in scoring marks for relative degrees of 'poor work.' The tradition interferes with the plasticity of mind which an examiner needs. Ques-

tions vary quite unexpectedly in the difficulties they present to the candidates, yet the marks recorded per question should approximate to the normal distribution, *i.e.* the same proportion of candidates should score 7 to 10 out of 10 on all those questions, which have been set with the intention that they should present equal difficulties. The tradition prevents examiners from scaling up the marks for questions which prove unexpectedly hard, and from scaling down the marks for questions which prove unexpectedly easy.

VI. Examiners should be specialists. The tradition that specialists make the best examiners has this to recommend it, that they should be specialists in examining. Specialists in separate subjects are less likely to make good examiners. Brilliant mathematicians with a penchant for strictly logical analysis and expression, or English scholars with a highly developed taste in regard to correctness of diction, are, as a rule, too hypercritical for the task of examining the work of children of ten or eleven.

II. THE BREAK-AWAY FROM TRADITION

Wholesale Condemnation. Critics of examining and experimenters in methods of testing intelligence have during recent years been faced with a dilemma. The traditional examining system has, on the whole, and in the long run, worked fairly well; in most cases the examiner's verdict has been supported by the subsequent career of the candidate. Yet almost every experimental test of examining methods has seemed to point to a weakness in the system. For example, half of a set of experienced examiners would have failed a candidate for a piece of Latin prose which the other examiners would have passed.

Consequently, the new examiner condemns the old system, alleging that the successful results of the traditional system were purely accidental, and that the methods did not necessarily achieve the desired end; they were lacking in scientific precision.

Dr. P. B. Ballard, in *The New Examiner*, pp. 57-61, quotes a case which is likely to be referred to as a classic instance of the failure of the traditional system.

K. B. was a little girl of nine and a half years who sat at an examination in which she was competing against her classmates for a qualification to sit at a

later competitive examination for a scholarship to a well-known public school for girls. She 'was required to write a story to illustrate the proverb: One good turn deserves another. . . . The examination essay bore few signs of exceptional ability.' The next day K. B. wrote at leisure on the same topic, and produced an essay which Dr. Ballard rightly quotes *in extenso* as 'a remarkable piece of writing for a child of nine.' Dr. Ballard quotes a set of verses by K. B., and mentions facts about K. B.'s record in regard to various mental tests. On the evidence he concludes that K. B. 'does not shine at examinations, and, indeed, it is doubtful whether any of the accepted types of examinations could ever reveal the full flower of her mind.' 'This little girl has the intellect of an adult, and every type of new test to which she has been subjected (and she has been subjected to a large number) brings this fact clearly to light.' When Dr. Ballard wrote *The New Examiner*, the complete story of K. B.'s efforts at scholarship examinations could not be written, because K. B. did not take her real test until after Dr. Ballard's book was published. The case may be considered in detail. At the qualifying examination to which alone Dr. Ballard refers, K. B. gained what she sought—the right to sit some months later at the subsequent examination. At the age of nine years ten months K. B. competed against four thousand other girls, most of them a year older than she, at a scholarship examination in which she was trying to secure a place in the top twenty. K. B. was then an under-age candidate, put forward on the ground

of her exceptional ability. She was examined in Arithmetic and English, and her work was of sufficiently outstanding merit for inclusion in the Borderland. She did not gain the scholarship she aimed at.

A year later, at the normal age, K. B. competed again, and this time gained her place in the top twenty by a very comfortable margin.

These details indicate only one fact, that K. B. is representative of the super-normality which characterises the most brilliant children of London. The London scholarship system discovers such children twice a year. It may be added that K. B. was recorded as a number, and that none of the facts here stated was known until after the final success had been achieved. The complete break-away from tradition of the new examiner has apparently followed a wholesale condemnation of the old system, and it may be well to pause for a moment to consider what such a *volte-face* implies.

Fear of the results of human frailty. Just as the keynote of the secrecy of the old system was fear, so the new examiner is afraid—afraid of the consequences of human frailty. Because an examiner, being human, is incapable of the precision of a machine, the new examiner wishes to eliminate from examining the judgment or appreciation or assessing power—call it what you will—of the examiner. Experience over a period of twenty years shows that the frailty of the examiner is, at least, avoidable, because it is mainly due to thoroughly inadequate preparation. Arithmetic is supposed to be an easy subject to examine because of its apparent precision.

This is another of the incorrect examination traditions. Give six independent examiners the same set of papers in Arithmetic, and ask them to assess their value without giving them definite and precise instructions regarding the detailed valuing of the work, and they will in all probability thoroughly disagree. Bring the six examiners to a round-table conference, and let them agree upon a detailed schedule of marking, and they will assess the scripts with mechanical precision. Arithmetic is no easier to mark precisely than English or History; there are just as many opportunities for divergence of judgment by examiners in Arithmetic or Latin or Mathematics as in History, Geography, or an English essay.

The new examiner favours fragmentary mental tests and numerous tiny exercises. Because experiment has demonstrated a possible, and perchance probable, instability of examining of lengthy and complicated tests, it has apparently been concluded that such instability is inevitable, and, hence, that such lengthy tests are unmarkable, and, therefore, useless. The determining factor is the markability of the answers.

It may be that the investigator into methods of examining when faced with the barriers of tradition mentioned in the first chapter, has lacked the material from which to discover whether the faults to which he has directed attention are accidental to defective machinery or are necessary consequences of a system. In any event, it would seem that a system which has been, on the whole, and in the

long run, successful has certain merits as well as obvious defects, and it would be desirable to know whether these merits and defects are accidental to the personality and ability of certain capable examiners, or necessary results of a system which seems effete and outworn.

It would not appear that either the old or the new examiner has implicit confidence in the ability of trained examiners to conduct an examination test upon the traditional English lines.

Fool-proof methods. The primary consequence of this dread is the advocacy of fool-proof methods, of the elimination of the examiner's judgment. Every examination answer must be such that it can be marked right or wrong, and every examination script must be such that the actual marking can be performed mechanically by a person who, on occasion, may have no knowledge of the subject examined. For example, a paper in Latin might be set, and the actual marking of the scripts might consist of the comparison of a series of small answers with a schedule. Each answer is right or wrong, and a careful person ignorant of Latin might mark the scripts satisfactorily. Theoretically the valuing of the candidate's work is perfectly consistent; it is claimed to be fool-proof.

Losses and gains. Superficially the gain from such fool-proof methods is immense; the variable and, possibly, inconsistent assistant examiner is eliminated. But the losses are, at least, equally tremendous. Everything is prepared for the candidate; he faces one issue at a time; he is never

tested on his powers of selection, of marshalling the data. The examiner does most of the thinking ; the candidate is not required to give reasons for his answer, although there is some ability indicated by a train of argument which fails in one step only out of four or five. Very frequently it is not the correct answer which matters, for the argument on which the answer is based is the test of intelligence. If a thousand or so answers to a problem in Arithmetic are examined it usually happens that the answers fall into five categories—one quite correct, another hopelessly wrong, and three others partially correct. These represent five grades of intelligence in the face of the problem set, and fool-proof methods fail to detect these grades. Further, short answers of a word or two give the candidate no chance to display ability in presenting a case. It is a legitimate test to state a problem and its solution, and to require the candidate to demonstrate the accuracy of the solution. In some cases such a demonstration indicates that the candidate would probably have solved the problem successfully ; in other cases it indicates little more than an intelligent grasp of the whole, *i.e.* the problem and its solution. Fool-proof methods fail to test deductive ability of this character.

One other objection is important. Intelligence is at times manifested by constructive ability or by ability to carry through to a successful issue a complicated operation. The power to weave a pattern, either with or without materials supplied, is an attribute of intelligence, and this power should be tested.

Essays. In this connexion the essay or composition exercise is of prime importance, although the essay is not the only material for a test of constructive ability. The new examiner has no faith in the ability of any set of examiners to mark essays consistently. The old examiner has left on record no technique for marking essays satisfactorily. Investigation has, however, elicited the following. With careful preliminary treatment, so that the examiners are agreed upon what they are going to mark, and with experienced examiners who are in preliminary agreement, a system of consistent essay marking can be established. It is demonstrable that with such a system a thousand candidates can be 'strung along a line' in order of merit at least as successfully by means of essays as by any form of examination test. This matter may be left here; the details upon which the statement is based will appear in a subsequent chapter.

Mental age. A useful conception upon which the new examiner lays great stress is that of mental age. The child of great intelligence is indicated when he passes the tests normally passed by an average child of a year or two years' greater age. The very intelligent child of eleven might even be credited with the 'mental age' of an average child of fifteen. It is worth while to envisage for a moment what this conception implies. It suggests that the child of eleven is as capable as the child of fifteen, but this is too sweeping a suggestion; in fact, it means that the child of eleven will score as many marks upon certain mental tests as the average child of fifteen

This fact, however, ignores one possibility. The child of eleven has not the experience of life and the world of a child of fifteen, and it would be possible to arrange the content of a mental test upon the basis of this fact so that the child of eleven could not score the average marks of fifteen-year-old children. This fact was suggested in connexion with the case of K. B. The essay produced in the first examination when K. B. was under-age was relatively immature in comparison with the more polished performances of the best of her competitors.

The content of an examination test. It is a commonplace in examining that the form, the matter, or the content, of the test is an essential part of its difficulty. Two questions in Arithmetic, for example, may have the same intrinsic difficulty as an exercise in thinking, but the form in which the problem is presented, the words and quantities used, *i.e.* the content of the question, may make of the one a question suited to children of eleven, and of the other a question for children of fifteen which children of eleven would find too difficult. Even the most experienced of examiners occasionally sets a question of which the content presents unexpected difficulties to the candidates.

The completeness of the break-away. Essays and essay-type examination papers are barred. All questions hinge upon one mental process, one bit of thinking, which results in a solution which is assessed as right or wrong. The judgment of the examiner is eliminated. Constructive power is not tested. Insight into a complicated problem is tested but

slightly. The candidate's ability to 'state a case' is not called into question. The content of the test must be within the experience of all the candidates. Alternative methods of solution of varying degrees of intelligence are not desired. Partially correct solutions receive no credit. The child who fails at the whole is marked no lower than the child who fails at a part.

The break-away from tradition is almost complete, with the result that examining becomes a much more limited operation than of yore. It is, perhaps, of some significance that the main impetus towards mental testing comes from America, and that the new examiner is, perhaps typically, an American 'mass-production' expert.

III. THE MODERN EXAMINER

The compromise. The modern examiner is condemned over and over again in actual practice to a compromise. He starts with a compromise, by an attempt to make the most out of the methods of both the old and the new examiner, and throughout his work he compromises between what he would like to do and what he is able to do. Being English, and therefore, if you like, insular and conservative, he cherishes the old system, but being modern and a pioneer, he is keen to get the best out of the new system.

Unlike his predecessors, he has no fear of human frailty. He recognises its existence, and lays his plans to limit and control the human examiner so as to make use of his judgment, and yet be sure that he is consistent in valuing the candidate's work. He takes a very limited view of examining ; his sole business is to string the candidates along a line in an order of merit.

The objective method. The modern examiner has no prejudices ; he is not concerned with what candidates ought to do or ought to know. A batch of candidates is represented by a number of scripts ; his only task is to arrange these in order of merit. He

aims at a list in which 50 per cent. of the candidates score at least 50 per cent. of the marks, and in which there is a reasonable proportion of candidates with scores between 90 per cent. and 100 per cent., and an equal number with scores between 0 per cent. and 10 per cent; he aims at a normal distribution. Theoretically he is prepared to sacrifice many things to secure this end; in practice he compromises as usual.

He is not concerned with the suggestion that the average candidate who scores half marks ought to have scored 60 per cent., or with a comparison between the average candidate of this year and the average candidate of previous years; he has no concern with the decision where the pass line should be drawn. It is immaterial to him whether the top 30 per cent. of the candidates are to receive scholarships, or whether the lower 35 per cent. of the candidates are to fail and be debarred from certificates. It does not matter to him whether the scripts represent bad or good teaching, a high or a low degree of attainment. He takes the scripts as he finds them, and adjusts his methods in such a fashion as to produce the desired order of merit list. He is inhuman, dispassionate, unbiassed; he tries to weigh and assess in scientific fashion the relative merits manifest in the batch of scripts under examination. All other ideas are adventitious and accidental, quite foreign to his impersonal and objective outlook.

The modern examiner thus breaks free from the traditional trammels of the old system without accepting more than the spirit of the new system.

His methods are not fool-proof; he admits the possibility of error, but limits its likelihood; he is prepared to measure the degree of error and make the necessary adjustments. He has no fear of the results of his methods as applied to any examination and to any subject. He makes but one demand —sufficient time to make an adequate preliminary analysis of the scripts, in addition to the usual time allowed for their valuation.

Divisions of an examiner's task. In connection with an examination there are, usually, a syllabus expressed or implied, a set of question papers, the candidates' scripts, and the final list of awards whether scholarships or certificates. The foregoing paragraph has dealt solely with the candidates' scripts, and these are primarily the assistant examiner's concern. Others may decide on the syllabus, others may set the papers, and others may authorise the awards; all these others reach their decisions for a multitude of reasons, many of which are outside his province. As examiner, his concern is the candidate's work and the order of merit list containing the labels (not the names) of all the candidates.

As the result of his main work a chief examiner may be called upon to advise upon a syllabus, or upon a published pass list, or he may be required to submit suggested papers of questions which are approved and authorised by others. In that event a chief examiner may influence the whole of an examination, but such an influence is additional to his principal task.

The best examiners. It follows from the foregoing that the examiner should be an expert not in a subject, but in examining. No class of trained persons necessarily provides the best examiners; the desirable powers of appreciation, of close application and concentration, of trained memory, of sustained effort, can only be discovered by experience. Examining is an exacting business, carried through at high pressure during a limited time. There is an old tradition that examiners should be changed frequently, that a spell of three to five years is long enough for an examiner to grow stale and to have given his best.

Provided there are sufficient expert examiners to carry out all the examining there is to be done, the tradition may be wise, but, in fact, it is doubtful whether there are sufficient expert examiners. The moment the tradition that a good teacher is probably a good examiner is discarded, there is a limited field from which the supply of examiners may be drawn. Examining further suffers from another disability. Usually examining work is an extra, a special job undertaken for the sake of its financial return in addition to one's customary labours. There are few examiners who are paid sufficient for them to be examiners and nothing else; most examiners have other interests; to a very high proportion examining is of secondary, or possibly of casual importance. They lack the necessary opportunities or incentives to become experts. It is noteworthy under these circumstances that examining in the past has been so well done.

The preliminary survey. Before closing this chapter, which completes the preliminary survey of the ground under consideration, it is desirable to point out that the ground has only been cursorily covered, that only a few main notions have been presented for consideration, with a view to preparing for the more detailed discussion of the technique and art of examining which follows. The old system has only been seriously called in question within the present century ; the new system has flourished in America for but a few years, and has only recently been adopted in England. Consequently the system to which the label modern has been attached for convenience has no past history to justify it ; in fact, the technique of the modern examiner is only in process of determination. The modern examiner has no fear of publicity ; on the contrary, his technique will be improved by enlightened criticism, hence what the traditional examiner might condemn as a lack of reserve characterises the chapters which follow.

IV. FUNDAMENTAL ASSUMPTIONS

The examinations. Children's examinations are held for the purpose of awarding certificates or scholarships. Scholarship examinations range children, of ages which vary from ten to twelve, according to those qualities to which the terms intelligence and ability are frequently applied, the qualities which are here labelled by the term capacity. Successful candidates proceed to secondary schools, usually for a term of four years. Most of the candidates are from primary schools. The principal aim of these examinations prevents their use as tests of attainment, and the subjects of examination are as general as possible in their scope, and are the native language and elementary ideas of quantity, usually designated, for convenience, as English and Arithmetic.

Certificate examinations are more complicated. The candidates are fifteen or sixteen years old, and have concluded a four-years' course at a secondary school. The successful candidate receives a certificate which serves a triple purpose. It indicates an average measure of capacity, it denotes that the candidate has reached a certain level of attainment in specified school subjects, and, with definite

limitations, it is a matriculation, and qualifies the candidate to proceed to a degree course at a university.

In these pages the second and third functions of certificates may be taken for granted, since it may reasonably be assumed that a child of average capacity should attain the levels of attainment indicated by these other functions. In a purely general discussion this assumption must be justified. If a school is defective and the average candidates from that school fail to gain certificates, the fault lies in the school. The few candidates who require the matriculation qualification may be left to the regulations of each university. In neither case do these particular cases derogate from the value of the general assumption—that the prime function of a certificate examination is a test of capacity.

The fundamental hypothesis. The modern examiner relies entirely upon an hypothesis which he cannot prove. Given thousands of candidates drawn year after year from the same sources at the same ages, under, roughly, constant conditions of entry, the standard of capacity of the average candidate will be either constant or subject to smaller secular variations which may be ignored.

Conditions of constancy. Average capacity will be constant from year to year. The form of manifestation of capacity may change, but there is no reason to suppose that the average candidate of 1926 is more or less capable than the average candidate of 1916.

Provided groups of candidates represent similar

groups of the population, their average capacity will be constant. It does not follow necessarily that the average capacity for Lancashire or London scholarships, or for Oxford or Cambridge certificates, is the same.

Provided the examinations are sufficiently general in their scope the average capacity they indicate is constant; the same candidates should score half marks.

The net result of these considerations is that not only the examinations of each authority, but even the examinations from year to year, form individual problems which the examiner is set to solve. The necessary compromises which occur in practice give to each separate examination its individual character.

Why take great pains over examination results? From the considerations just enunciated it might be argued that the right candidates come to the top under any examination. This, with some reserve, may be admitted. The examiner is not troubled with the top candidates; he is troubled, however, with border-line candidates, *i.e.* with all the candidates who are within an ace of success or failure. A scholarship or certificate may be of enormous value to an individual candidate, and the examiner is justified in leaving no stone unturned to secure within the limits of his authority that no candidate misses a scholarship by any fault of any description in the examination system.

He must feel sure that, so far as he can secure precision and accuracy, the individuals who have

secured awards are the right individuals. His records must be such that they will stand scrutiny ten years hence.

There may be other reasons which suggest that the precision at which the examiner aims is negatived by circumstances which he cannot control. These reasons do not justify failure to perfect the machinery of examinations.

A working hypothesis. Among thousands of candidates capacity may be graded numerically by marks out of a maximum of 100 so that the distribution of marks will follow what is called the normal curve of distribution. This normal distribution always occurs with an unbiassed investigation into any common property of large numbers of individuals. The examiner uses the principle of normal distribution as a working hypothesis.

The marks scored for each paper should be normally distributed, *i.e.* half the candidates should score at least half marks, and as many candidates should score at least 75 per cent. as score no more than 25 per cent.

The degree in which the candidates are bunched close to, or spread far from, the 50 per cent. mark is entirely within the control of the examiner, and he manages this dispersion of the candidates for the separate papers and for the examination as a whole so that he may secure a reasonable spread of the 'border-line' candidates. He adjusts his methods in detail to each individual examination.

It has been implied that the dispersion is controllable. This notion requires some expansion.

There are two methods of control. The first is mathematical, and consists in a mathematical adjustment of the marks recorded by the examiners. The dispersion due to the original marking of the scripts is allowed to take its chance. It may not even be quite normal; it is modified until it is normal and of the form desired.

Experience shows, however, that the original marking can be so arranged as to give a normal curve of distribution with any desired dispersion. It has been thought that essays and essay-type papers give a different form of dispersion from arithmetic-type papers. This is not fundamental. A schedule of marking for Arithmetic can be arranged to give a dispersion such as has been habitually obtained for essays; and similarly essays may be so marked as to give a dispersion of the type habitually associated with Arithmetic.

The expert modern examiner is able to secure any dispersion which he desires, and is able to adjust his technique to the particular task before him. This is the second, and the better, method of controlling dispersion. Apart from theoretical considerations, it has the practical advantage of saving time. But this is a merit of minor importance in comparison with the added flexibility which this method of control gives to the examiner's technique. He may design his papers and plan his marking with the promise that when the marking is done he will have secured a normal distribution and a suitable dispersion. This circumstance is one measure of justification for the suggestion that the examiner

should be an expert in examining rather than an amateur who examines for a year or two for the sole purpose of adding to an exiguous income.

Practical consequences. Following upon the assumptions made in this chapter, there are certain important practical consequences. Each scholarship examination is an individual problem for the examiner. The methods suitable for the award of six hundred scholarships when six thousand candidates are examined are different from those suitable for an award of one thousand six hundred scholarships when six thousand candidates are presented. The methods to adopt for candidates from Pembrokeshire differ from those suitable for Northumberland or Sussex. Hence the examiner requires time to adjust his technique to the local individual conditions, and, further, he should develop a suitable technique which he can pass to his successor. Examiners should act for a longish term, and should, on laying down their charge, leave behind a precise and detailed statement of the methods they used. Too frequently an examiner on taking up an appointment has to buy his experience, and to give up his appointment before his experience has been of much use in the adjustment of his general methods to the individual problem.

A certificate examination should result in, roughly, a constant percentage of certificates to candidates, and, further, a roughly constant percentage of passes per subject and per annum.

Granted that capacity is the main quality under examination even for school certificates, and granted

that average capacity is on the whole constant for thousands of candidates, the authorities who control certificate examinations have the duty to prescribe beforehand the percentage of candidates who shall pass on each paper and who shall gain certificates. This is not the usual practice, hence the anomalies that 60 per cent. of the candidates pass in a subject one year and 40 per cent. the next year, or that 60 per cent. pass in subject A and 35 per cent. in subject B; while the percentage of certificates gained varies from year to year.

It is contended that the usual practice of leaving the percentage of passes to chance is entirely wrong. The customary procedure throws a responsibility upon the examiner which is not rightly his. The examiner has no data upon which he can safely conclude that the standard for a pass in his subject is the right standard, and that the standard is steady from year to year. He may have faith that his standards are right; his critics aver that his standards are neither right nor steady; and the balance of probability is in favour of the critics. The matter has not been dispassionately investigated, and the safer practical plan is to determine approximately the percentage of passes beforehand, and to require the examiner when he reviews the 'border-line' cases to report each year upon the merits of the 'border-line' candidates in relation to what he assumes each year to be the requisite standard for a pass.

An examiner can recognise a certain pass paper or a certain failure, *i.e.* papers well above or well

below the average, but he is quite incapable of recognising a paper which is a bare pass. It is not usually possible for an examiner to justify the award of 49 marks to a script when 50 is the pass mark. It is asking a great deal too much of an examiner who is working at high pressure over a period of several weeks to suggest that he is capable of so fine a distinction.

It is grossly unfair both to an examiner and to the candidates to compel him to mark the scripts when he knows that a 49 will fail and a 50 will pass. It is fair to compel him to mark with especial care the scripts which total between 45 and 55, because some of these will fail and some will pass; but the final decision should not rest with him. The authority should say, for example, 55 per cent. of the candidates should pass in English. The examiner should then consider as 'border-line' cases the last 300 or so of the 55 per cent. or more group and the next 300 or so below the 55 per cent., and after a careful scrutiny of the scripts he should separate the sheep from the goats, and recommend, perchance, that the percentage of passes should be for this particular year 56 per cent., or even 53 per cent.

The disgruntled candidates or their teacher would then know precisely where they stood. Seven thousand candidates took the same paper; their scripts were marked consistently, and only the top 55 per cent. or so passed. The capacity of the candidates who failed was definitely registered as below the average. They definitely did not produce evidence that they were as good as the normal pass

candidate in this paper at this examination. It would not be possible for a candidate to excuse his failure by saying, 'Oh, they ploughed an unusual number this year,' or, 'The papers were stiff, and so there was a wholesale slaughter of the candidates.' He could not even justifiably suggest, 'There must have been a good lot of candidates this time, and the standard was high.' So far as can be judged, the standard would be definitely the same from year to year, and would be independent of changes in syllabus, differences of difficulty in the paper, changes in examiners, *i.e.* of any of the numerous variable factors which affect examinations.

Finally, it may be suggested that a practical consequence of these hypotheses is the release of the examiner from the trammels of the schools. It is not the examiner's business to do more than solve a particular problem annually presented to him. It is not his job to criticise or improve the teaching; he is not called upon to be a pioneer in education; he should be a rather conservative, open-minded spectator. His place is not in the vanguard of educational progress. He is not concerned with the *media* in which capacity is expressed; his business is the quest for capacity and its assessment upon a scale of relative values.

Put the matter in an extremely blunt way. If all the teachers idled for a year, it is not his business to plough their candidates; it is his business to take the scripts as he finds them and grade them in order.

The examiner for scholarships for secondary schools is not called upon to say how many candi-

dates are fit for secondary education. The authority determines the number of scholarships, and the examiner says which of certain individuals shall be included in that number. The authority for certificates should say so many candidates shall pass in this subject, and the examiner merely provides a list of that number of individuals.

This practical consequence gives point to the suggestion that the expert examiner need never have been a teacher, and need not be familiar with the schools whence the candidates are presented.

V. SYLLABUSES

Preliminary notions. For scholarship and certificate examinations a syllabus is either implied or prescribed. In both cases the responsibility for the syllabus rests upon the authority. The examiner may be an expert consultant; his reports to the authority may influence the syllabus; but the examiner is strictly without power to lay down the syllabus. The authority determines the purpose of the examination, and, hence, the syllabus on which the examination papers are to be set.

A syllabus should be general, definite, and easy. It should not imply a definite standard of attainment except for specific and specified purposes.

A syllabus for an examination which attracts thousands of candidates from hundreds of schools should be sufficiently general, so that no school nor type of school is specially favoured. A syllabus loses in efficiency the more it enters into details; it defeats the object of the examination—the quest for capacity—the more limitations it imposes. Capacity is demonstrated differently in different settings, and the syllabus determines a setting in which capacity is to be discovered. The more limited the setting the more difficult is it for capacity to become manifest.

The syllabus is the first way in which the examination affects the schools, and no prescribed syllabus is infinitely to be preferred to one which is ambiguous. A syllabus should leave no doubt as to the intentions of the examining authority.

The syllabus should prescribe for the candidate who is slightly below the average. Examining experience indicates that it is easier to detect capacity with some measure of certainty on an easy syllabus than on a difficult one. The able candidate should feel that he has mastered the whole of the prescribed syllabus. Hence the syllabus should not be too long nor too complicated.

Scholarship syllabuses. It is customary to prescribe for scholarship examinations by stating that there will be papers in Arithmetic and English. This prescription should imply Arithmetic (or English) to the level of attainment of approximately the average candidate submitted by the schools. This implication implicitly bars formal grammar, except possibly in its very rudiments, from the English paper, and the more advanced weights and measures from the paper in Arithmetic.

It becomes, therefore, the business of the authority to determine and place on record (on record confidentially, if secrecy is desired) the upper limits of the knowledge which is implied by the words English and Arithmetic.

For example, each authority has to decide whether knowledge of volumes and cubic measure, knowledge of percentages, ready facility with decimals, fall within the syllabus in Arithmetic.

An Advisory Board. In practice probably the best means of securing a satisfactory syllabus is for the authority which is necessarily busied with less technical matters to delegate its powers to an Advisory Board. This Board should be representative of the officials of the authority, the teachers of the primary schools which submit candidates, the teachers of the secondary schools to which the successful candidates go, and the examiners.

From the examiner's point of view the Advisory Board should be definitely weighted in practice to give great effect to the opinions of the teachers of the primary schools; in fact, the representative teachers in primary schools should chiefly represent those schools which are of such a low level in the capacity of their pupils that they secure a scholarship award only occasionally. They should represent something lower than the capacity of the bottom scholarship child.

The Board should be thoroughly alive to its main responsibility—that no capable candidate suffers disability because the syllabus represents a standard of attainment which he has not normally reached.

The syllabus and the secondary school. The general scholarship examination which applies for a group of secondary schools is a different proposition from the admission scholarship examination of an individual secondary school. In general the authority has decided to give so many children the chance of a secondary education; in particular the Head of the secondary school wants to know where and how a particular child will dovetail into his school. In

some areas the successful candidate at the general examination undergoes further examination for one or more particular schools.

In the general case the educable capacity of the candidate is rated in relation to his age group, and the scholarship examination at which the child has been successful does not specify that the child has reached more than a limited standard of attainment, usually in Arithmetic and English and in no other subjects. In the particular case educable capacity is taken for granted, and the Head of the individual secondary school lays down an extensive syllabus which should refer to a greater degree of attainment in Arithmetic and English, and, possibly, in History, Geography, Elementary Mathematics, and French than any but the exceptionally endowed scholarship child can reach. Obviously, if a Head has twenty vacancies for successful scholarship children, and has usually applications from forty or more children to fill these vacancies, his particular admission examination will resolve itself into a competition in attainment and in such factors as zest for games, etc., and his particular syllabus will aim at filling his vacancies with those children who suit his school best. If, however, there be twenty vacancies and normally fewer than twenty applicants, the Head may prescribe a syllabus merely in order to grade his candidates by attainment and determine the forms to which they shall go.

School certificate syllabuses. The syllabus for each subject in a school certificate examination suffers from a drawback, in that it is governed in

part by the circumstance that it is a qualification for matriculation. The university authorities demand a minimum of attainment for matriculation, and the school certificate syllabus must embody this minimum.

In practice the general certificate syllabus contains such a minimum, together with suggestions of sufficient material to cover a course in the subject at a normally good secondary school. Many teachers find that they can base their course on the syllabus so that their children can take the certificate examination in their stride as they proceed through the school. To meet the needs of other teachers, opportunity is usually provided for candidates from an individual school to be examined upon a special syllabus which has been approved by the authority. In such cases the approval of the authority will hinge upon the fact that the special syllabus on the whole is a fair equivalent of the general syllabus.

Options. In practice it is very difficult to make options which are relatively equal. For example, in Geography a general syllabus might prescribe :

> a knowledge of geographical principles as applied to the world as a whole ; a detailed knowledge of the British Isles ; and a less detailed knowledge of the rest of the world, with decreasing detail as distance from Britain increases.

In place of the third item it might be made optional to substitute :

> a general knowledge of (a) Europe or (b) North America or (c) the Southern Hemisphere.

Obviously these three options are not quite equivalents either to each other or to a 'less detailed knowledge of the rest of the world.'

Options, also, tend to obscure the main purpose of the school certificate examination, which is the quest for educable capacity which has been educated to, at least, a definite minimum of attainment. Such a purpose is general in aim, and options obscure the general aim among particular details.

Special scholarship examinations. In addition to the examinations for candidates of the age 10+, some authorities examine older children for scholarships or admissions to secondary, trade, and other schools with a view to continued education. It is a practice that these children shall be examined in Arithmetic and English and one or more of the following subjects: History, Geography, Needlework (girls only), Geometrical Drawing and Woodwork (boys only). Obviously these additional subjects involve an examination in attainment, and a syllabus may be prescribed.

A syllabus is desirable, whether it be prescribed and published or determined and kept as a confidential document for official use. Such a syllabus should be general, of wide application, and should follow the school practice of the locality. It should be known within wide limits to the authority and to the Advisory Examining Board how much information about History or Geography, for example, the average examination candidate might reasonably be expected to have assimilated. It would be persistently unfair to batch after batch

of candidates to use the Geography in such an examination in order to improve defects in the teaching of Geography in the elementary schools from which the candidates were submitted to examination.

VI. EXAMINATION PAPERS.

Responsibility. The responsibility for the examination papers rests upon the authority or upon the Advisory Board as representing the authority. Neither chief nor assistant examiners should be solely responsible. In practice, examination papers are suggested by an examiner, and are modified, adapted, adopted, and approved by the authority.

This practice is an advisable compromise. The examiner who suggests the paper utilises his experience to provide questions which are a test of intelligence, and which will produce markable answers. As an examiner his main concern is the provision of questions which will assist him to classify the candidates in an order of merit. Within the syllabus he has no recognised limitations.

The authority, however, must consider his suggestions from many points of view. Are they within the syllabus ? Will every child get a fair chance ? Will the effect of a question or of the whole paper be a harmful reaction in the schools ? Can the questions be defended or justified in the event of hostile criticism by parents or teachers ? The net result is that an examiner who wishes to modify the type of paper in use at an examination must always

accept a compromise, must proceed by slow steps and 'educate his public.'

Practicability. A set of examination papers always compromises between what the examiner and the authority desire and what is practicable. The age and physical capacities of the children, the distances they would travel between home and the examination room, the time of day and the season of the year of the examination, the provision of examination accommodation and supervision, are all questions to be considered.

There is a known case, for example, when an examination was largely nullified by the circumstance that the candidates sat in pairs in dual desks in the examination room. The examiners found an excessive number of cases of collusion between candidates whose examination numbers were x (an odd number) and $x+1$.

The examiner might suggest an examination extending over two hours on each of five days in one week. He would probably be told that such a course was impracticable. He might suggest attendance at a twenty minutes' exhibit of a motion picture. He would probably be told that this would cost too much. He might desire the provision for each candidate of a definite solid object, or the demonstration in each examination room of a simple experiment, or the use by each candidate of a book or piece of apparatus—all which possibilities would be probably refused.

Each examination occupies the number of hours, and consists of as many question papers and as

many questions per paper as the authority deems advisable or possible, and the examiner is constrained to keep within these limits. Most scholarship examinations, to their detriment, involve merely the provision of printed question papers and answer books ; they are too ' bookish.'

The Advisory Board. The composition of the Advisory Board is closely related to the compromise which every set of papers puts into effect. The purely administrative matters—provision of rooms, desks, paper, superintendents, etc.—are represented by the officials of the authority ; the restraint due to the limited attainments of the candidates receives special consideration from the representatives of the schools which send them ; the ideals underlying the future education of the children guide the advice of the representatives of the secondary schools ; and the views of the examiners are duly supported by the examiner members of the Board. A composite body of this character naturally produces examination papers which are a compromise. In one particular aspect a section of the Board should have what is tantamount to a veto upon individual questions ; the representatives of the elementary schools should say definitely regarding each proposed question whether, in their opinion, the children can be reasonably expected to answer it. If the opinion of the teachers is negative, then the question should not be approved unless it is to be included for a specific reason.

A chief examiner for scholarships. As a consultant expert the chief examiner for scholarships may

suggest the questions and exert a guiding influence in the form of the papers. His influence is in many cases a decisive factor in the result. His reports, to which detailed reference is made in a later chapter, upon the last examination necessarily affect the preparations for the next examination.

Two of the chief examiner's prime duties are concerned with the effect of the separate papers upon the examination as a whole, and with the whole performance of the candidates, the more especially in the case of candidates in the borderland between success and failure. Of the individuals concerned in the examination, he alone is in a position to regard the work as a whole, to co-ordinate the different elements, and to preserve a just balance between varying circumstances.

The provision of suggested examination papers is, therefore, the chief examiner's chief care, for it is the principal way in which he can bring his special activities to bear upon the work. His position should be a strong one; his suggested questions and papers as a whole should be entitled to the approval of the Board, unless the Board finds other than examination reasons for their inclusion. A suggested question sometimes gives rise to several discordant views upheld by different members of the Board for different reasons. In practice great discordance is evidence of an unsuitable question which should be withdrawn. Occasionally the Board is sharply divided into two camps; in that case the majority vote is usually an adequate index of a question's suitability.

Some Advisory Boards consider the papers suggested by the chief examiner in great detail; they weigh their merits sometimes almost word by word. Other Boards accept the papers almost exactly as suggested, with possibly the deletion of an occasional question or a request for an alternative. Other Boards leave the papers entirely to the chief examiner. Experience of all three methods suggests that the wisest plan is the detailed discussion of the papers question by question.

School certificate papers and chief examiners. In school certificate examinations there is usually a chief examiner (or examiners) in charge of a team of assistant examiners for each subject of the examination. Above all is an Advisory Board representing the authority. The Board selects the person to set the paper. The setter may be an outside expert, a chief or an assistant examiner. Practices vary even under the same Board in the same year. The Board and the setter determine the paper. The chief examiner may have some influence on the paper either as a setter or as a member of the Board, or as a specially consulted expert, or he may not see the paper until some time after it has been determined. Owing to a variety of reasons the interval between the setting of the paper and the examination may be so long that the report upon an examination is received by the Board after the paper for the next examination has been approved.

These differences in practice suggest that there is no standard practice which is universally applicable or desirable.

Types of questions. It is traditional in examinations that answers shall be written in the form of essays, *i.e.* in a connected, continuous prose passage, wherever possible. In Arithmetic and Mathematics, and in languages other than English, the tradition does not hold ; but in History, Geography, English, and to a large degree in Science, it is an established convention that the answer should adopt the essay form ; and a candidate pressed for time who submits an answer in the form of elaborate and full notes feels it incumbent to apologise to the examiner for his departure from custom. This tradition influences Arithmetic answers, for some candidates feel that they must add words innumerable, to make assurance doubly sure, to an argument which is perfect by all the conventions of Arithmetic.

Like most of the other traditions discarded by the New Examiner, the tradition of the essay-form is harmful. Both Modern and New Examiners agree that the essay-form interferes with the markability of the answers, and both would agree that in marking (*e.g.*) History the examiner should not be constrained to mark English at the same time. The tradition interferes seriously in connexion with Geography. Many Modern Examiners would agree that some questions may be best answered completely and satisfactorily by means of a map in which are used all the requisite geographical conventions. Other examiners would accept a map with an addendum of verbal explanation. Such answers are extremely rare, and it is a commonplace in the reports of Geography examiners that sketch maps

are not used with sufficient freedom. The interference, however, goes further than this, since questions other than those of the essay type receive but a lukewarm approval from Advisory Boards. An ideal examination paper in Geography might not contain a single essay-type question, yet an Advisory Board would refuse to sanction it on the ground that the questions were too 'practical,' too 'scientific,' or too 'something or other,' and were, at the same time, too expensive.

The tradition also interferes with the inclusion in English papers of questions which may be labelled 'Intelligence-type' questions.

The fundamental feature of an intelligence or mental test as set by the new examiner is that it has already been used and tried, so that norms of performance have been registered, *e.g.* question A is correctly answered on the average by x per cent. of children aged nine and a half years. An 'intelligence-type' question adopts the form of question used by the mental tester, but has not been previously used. An 'intelligence-type' question requires that the candidate shall think and shall record the result of his thinking in a word or two. The tradition for 'essay' answers, for sentences which are really exercises in English composition, interferes with the use of 'intelligence-type' questions, and Advisory Boards are prone to sanction such questions but rarely.

An example of this sort of traditional interference may be recorded. A chief examiner was required to test the same children in both History and

Geography. Experience showed that previous papers had been too difficult for the candidates submitted, that it was impossible for the average candidate to score half marks. The examiner diagnosed the trouble as (i) immature candidates who were as yet not accustomed to think a great deal in terms of geographical and historical material; (ii) an insufficient attainment in either subject to permit the construction of a continuous narrative, so that 'essay' answers usually took the form of a reproduction of elaborate notes which had been memorised; (iii) considerable variation in ability to write pithy English about any subject. Based upon his diagnosis, the examiner introduced into both papers questions of the intelligence-test type. The result, from his point of view, was an immediate success. The average candidate scored very nearly half marks, and the candidates were well spread out, instead of being bunched excessively in the neighbourhood of 30 per cent. of the marks.

The Advisory Board sanctioned the first papers as an experiment, and, in consequence of its success, continued their sanction to subsequent papers. Experienced critics, however, who were acquainted with the circumstances, and had access to the scripts worked by the candidates, adopted a very curious attitude—they accepted the questions in Geography without comment, but objected to the intelligence-type questions in the History paper. Geography appears, therefore, to have escaped from some of the trammels of tradition.

From the examiner's point of view it is axiomatic

that the question set must be suited to the capacity of the candidates. Any question not so suitable represents a compromise between the examiner and other members of the Advisory Board.

Tests of attainment in Arithmetic. The tradition just mentioned affects Arithmetic papers in somewhat similar fashion. This may be illustrated in relation to an Arithmetic paper designed mainly to test attainment and, to some degree, speed in arithmetical processes on a syllabus well within the powers of the average candidate.

A paper was set containing five questions, each of which involved some written calculation. The amount of written computation sent in on the scripts depended upon two factors: (i) the ability of the child to perform computations mentally, and (ii) the child's training at school in setting out his arithmetical solutions. The quantity of written work submitted depended more upon the teacher than the child. Some candidates were so constrained by habit to a lengthy 'setting out' that they failed to finish the paper, and their scripts were endorsed by the examiner, 'slow worker.'

Yet under these circumstances too many candidates scored 'full marks,' and the number of questions was increased first to six and then to seven. Still too many candidates scored full marks, and the 'slow workers'—mostly 'slow' because of the type of answer they were trained to produce—were put at a greater disadvantage than ever. Success on this attainment paper was gradually becoming more and more due to the teachers than to the candidates,

and the examination was gradually failing in its primary object.

It was also demonstrable that the questions might accidentally involve a severe penalty upon an able candidate. For example, questions A and B each involve four steps. A 'slip' in the first step of question A might involve spending a third or more of the examination time in working out an answer by a laborious computation. An equivalent 'slip' or even the same slip ($3 \times 14 = 52$) in the first step of question B, or in the other steps of either question, might not involve additional computation. The examiner penalised each slip in accordance with his schedule, and the matter was closed for him, except that he was sometimes tempted to add a marginal note, 'Hard lines,' to the first case.

In view of these circumstances the type of paper was changed. The chief examiner made a new paper of twenty short questions, the answers to nearly all of which could be computed mentally. He took one of the old examination papers and extracted therefrom twenty computations which were involved in the old paper. The only change was to set each computation independently so that failure therein had no effect upon the rest of the work. It was merely a change of form, mainly designed to eliminate the effect of the teaching, and to examine the child and not the teacher. Still too many children scored 'very high' marks, although fewer children scored 'full' marks. The paper was made a little more difficult subsequently by the addition of a few questions involving more careful thought,

and the examiner was satisfied. He had evolved a paper which spread his candidates as desired, and gave no obviously special advantages to any child. His new paper then gave rise to the criticism that, while the old paper helped the teacher to train his children, the new paper reacted badly upon the schools. Tradition demands elaborate statement; the teacher must train his children to make complete and full computations on paper; and the examination will suffer because the examiner must bow to the tradition that the examination must lead the schools, not follow them.

Kinds of scholarship examination papers. It is demonstrable that there are six types of papers from which a scholarship examination can be built. Specimens of actual papers are in the Appendix.

I. Arithmetic Attainments: a paper designed to give the candidate who is 'quick' (*i.e.* lively, alert, etc.) at 'figures' a special chance.

II. Essay or Composition: an exercise in continuous prose narration, crystallised continuous speech.

III. English Questions: a paper meant to allow the candidate 'quick' at words a special chance.

IV. Arithmetic Problems: a paper which involves

(a) a comprehension of somewhat elaborate printed statements;

(b) considerable ability to think in rela-

tion to simple quantities expressed numerically or by diagrams ; and

(c) ability to reach a correct solution to the problem and to give some indication of the argument involved.

(The complete deductive demonstration of the solution cannot be legitimately expected from more than a few brilliant candidates.)

V. Intelligence Test (Verbal).

VI. Intelligence Test (Numerical).

These two papers may be well-tried mental tests or papers involving the intelligence type of question.

It is demonstrable that the essay paper and the two intelligence tests combined give results closely correlated to the general order of capacity determined by the whole of the papers taken together, and that the other three papers give opportunities for the manifestation of specific capacity of one kind or another.

The decision whether there be six papers or four, *i.e.* whether intelligence tests should be included, rests with the authority. A compromise must be affected between the desirable number of papers and the time allowed for the examination. Some authorities use intelligence tests for a preliminary sifting of the children in an age group to find those children who should be subsequently examined. Other authorities use intelligence tests only for candidates within the borderland. Yet others do not use intelligence tests at all.

The average arrangement of a junior scholarship examination might then be (i) Arithmetic attainments tested by intelligence-type questions; (ii) English questions with some intelligence-type questions in the paper; (iii) an essay or composition exercise; and (iv) Arithmetic problems which test English, Arithmetic, and 'intelligence,' as well as powers of inductive reasoning and deductive argument. To such an arrangement mental or intelligence tests would be supplementary.

School certificate papers. The difficulty of arranging that papers in twenty different subjects should be reasonably equivalent seems to be almost insurmountable.

Group.	Subject.	Number of Candidates.	Percentage of Passes	
			Year *X*.	Year *Y*.
I.	A	10,000	70	68
	B	10,000	76	91
	C	10,000	45	58
	D	10,000	51	56
II.	E	5,000	40	38
	F	5,000	47	51
	G	5,000	41	47
	H	4,000	54	27
III.	I	300	62	64
	J	600	39	10
	K	400	47	55

The above list is compiled from the actual returns of a school certificate examination solely to indicate the difficulties of the problem. Consider subjects

E, F, and G, for example, which attracted about the same number of candidates from the same schools with the same type of teachers. Assume that the main purpose of the examination is to discover capacity, the capability of the candidates in relation to different *media* of study and thought, *i.e.* mental ability in three different environments. Assume, further, that between year *X* and year *Y*, two successive recent years, there was no reason to suppose a marked difference in the average calibre of the candidates. In subjects E and F the percentage of passes remained roughly the same, yet one more candidate out of each ten passed in F than in E, and in G the percentage varied so that G was like E one year and like F the other. A form of thirty average pupils sat. It could be expected that eleven to twelve would pass in E, fourteen or fifteen in F, and twelve to fourteen in G.

It would appear that in school certificate papers a good deal is left to chance. For example, on subject B the examiners report that the paper in year *X* was 'slightly harder' than in the year *Y*, and the percentage of passes differed by 76 to 91. These same examiners give a clue to one of the chief difficulties of the situation. They say, 'The general impression left is that while the teaching of B in some schools is admirable, in many good, in a few it leaves something to be desired'—surely an unnecessary statement when 76 per cent. of the candidates passed.

The examiners in C whose passes were 45 per cent. in comparison with 58 per cent., explain that the general standard was hardly equal to that of recent

years, and that there was considerable difference (perhaps a greater difference than usual) between the work of different schools. The examiners in E, who had a 40 per cent. pass list in both years, state, 'As in former years the standard of work varied considerably between school and school. There was a large number of scripts of very poor quality, showing scanty knowledge of facts and mediocre powers of expression. On the other hand, there were groups of candidates manifesting accurate information, good reasoning ability, and first-rate power of expression. It is clear that the general average of passes can never become what it should be so long as the subject shows evidence of but perfunctory attention, as is the case in a number of schools.'

The examiners in E put up a poor defence of their 40 per cent. pass list. As regards mediocre power of expression, the examiners in English passed 60 per cent. Subject F requires a knowledge of facts, subject G requires the display of reasoning powers, as does subject C; and in the same year, among candidates from the same schools, obviously, in the mass, of the same range of capacity, subject E was a bad subject to take.

Without labouring these illustrations, it is perfectly obvious, *on the fundamental assumptions which have been previously stated*, that the school certificate examination is definitely trammelled by tradition. Some examiners are trying to lead the schools and to force the teachers to teach to their standards and ideals. On the face of it there seems little justifica-

tion for the charge of 'perfunctory attention' in these days of trained specialist teachers in each secondary school subject.

A solution of the difficulty. The examiners should be freed from any necessity to refer to standards of performance. It is not their business to say—something which is probably untrue—that the standard of performance this year is better or worse than or as high as that reached last year. It is the business of the authority to lay down *beforehand* a rule—the percentage of passes in subject A shall be x per cent. (approximately), unless for *very special reasons* the examiners recommend $(x \pm a)$ per cent. as a desirable standard for a given year.

Such a practice frees the examiner from a difficulty which frequently arises when the paper in one year proves more difficult than the average paper of previous years. The examiner is not required to equate the performance of a given year with the average performance of previous years. He can concentrate on his sole task—the equitable measurement of the relative capacities of a given batch of candidates.

Hard and easy papers. On general principles an easy paper is more suitable than a hard paper as a means of measuring relative capacity. A hard paper places too great a premium upon the candidate's good luck in having a teacher who manages to agree with the examiner.

But the test of the suitability of a paper has, in reality, nothing to do with its difficulty or its simplicity. A suitable paper produces the average

mark and the spread of candidates which in the scheme of the examination is desired by the examiner or *predetermined* by the authority.

If the first results of the paper do not produce the desired average and spread, then the marks should be adjusted accordingly; in scholarship examinations they are usually adjusted wherever necessary.

In some scholarship examinations the list in order of merit is divided into sections: (i) the top division, to which is attached the best type of award; (ii) a second division; and (iii) a third division, to each of which lower awards are attached; and (iv) the unsuccessful division which goes awardless.

In such cases the papers are a compromise, largely influenced by the size of the top division in relation to the number of candidates. If the top division is relatively very small, it is necessary to stretch the candidates to the limit of their capacity, and the papers must be such as will produce a reasonable spread of the candidates in the top division. It would be ludicrous if the top division contained thirty candidates, and in an extreme case the top thirty candidates all scored the same total mark. Consequently, the papers have to be arranged so that it is possible for only a few of the very ablest candidates to score full marks on a separate paper, and the questions in Arithmetic problems must present a severe test.

Zero marks. From the examiner's point of view as many candidates should score the zero as the maximum mark. It should not be regarded as impossible or unlikely that some candidates will fail

to score a single mark on a paper. In practice the papers of English questions and Arithmetic attainments are a compromise. They present a question or questions on which all candidates are expected to score some marks in order that the candidates may settle down comfortably to their work. The net result of this compromise is that the examiner's scale of mark distribution really begins at 10 per cent. of the maximum marks. Marks up to, say, 10 per cent. of the maximum on these papers, are a gift to almost every candidate.

Boys and girls. In scholarship and certificate examinations there is no competition between boys and girls, and there is no theoretical necessity that boys and girls should be set the same questions or the same papers. There is no doubt that boys and girls respond to the examination tests in a different way; they also respond to mental tests differently.

It was customary to state the difference in a crude way by saying that the boys did better in Arithmetic and the girls did better in English, and so the differences tended to cancel out. This crude statement is probably incorrect. It is demonstrable:

(i) that with the type of Arithmetic paper commonly set the boys score more freely than the girls, so that the average mark for boys is higher than that for girls.

(ii) that in essays and English questions there is no general or invariable difference.

(iii) therefore, that the differences do not cancel out.

It is probable that question papers in Arithmetic could be devised on which the girls would score at least as freely as the boys.

The setting of the same papers to boys and girls is, therefore, a practical expedient to be defended upon grounds of economy. Whenever the slightest chance of separate papers in Arithmetic arises the examiner asks for them. Obviously if a set of examination papers is rightly balanced to give the best possible order of merit list for boys, then it will not give the best possible order of merit list for girls.

In the case of school certificate examinations, there is much to be said for the suggestion that the authority should *predetermine* for Arithmetic and cognate subjects, such as Elementary Mathematics, that x per cent. boys should pass and that $(x+a)$ per cent. girls should pass. An elaborate investigation would be necessary to decide whether there are any examination subjects on which girls invariably score more freely than boys. The case in regard to Arithmetic papers of the class commonly set may be regarded as definitely established: the sexes perform differently.

VII. MARKING THE SCRIPTS

The examiner's task. In this chapter the examiner comes into prominence; in other chapters the assistant examiner appears incidentally, and the chief examiner as a consultant expert. When the time comes to mark the scripts submitted by the candidates the examiners put their principles to the test, and the chief examiner assumes a responsibility for the manner in which a definite task is performed.

Looking back upon early experiences of examining in the initial years of the century, one realises the chance way in which examining used to be conducted. One gained some reputation as a teacher or as an expert in connexion with a school subject, and one was invited to assist in an examination. Less frequently an examinership was advertised, and one became a successful applicant. Once appointed one received batches of scripts, sometimes accompanied by a schedule of marking. One marked the scripts as one thought best, wrote a report and returned the marked sheets and the report, and, sometimes, the scripts. If the appointment was continued, one presumed that the examining was satisfactory. Sometimes the schedule of marking was expounded, and, possibly, discussed at an

informal meeting of examiners. An assistant examiner received very few instructions, very little guidance, and decided most matters for himself and by himself. One absorbed from association with other examiners an idea that one's work was not, and should not be, questioned; one gave a verdict which was not modified or called into question.

In view of the fact that examiners, being human, make mistakes—mistakes of fact as well as errors of judgment—it seems rather extraordinary that the old examiner succeeded so well as he did. There was undoubtedly ground for the candidate's conviction that the results were often a mere fluke, that chance favoured some candidates, that failure was not always to be laid down to the candidate.

The cardinal principle. An examiner must mark scripts objectively. He is not concerned with what candidates *should* do, *can* do, or normally *do* do; he is solely concerned with the relative merits shown in the scripts he deals with, *i.e.* with what his particular batch of candidates *has* done.

He is not concerned with the quality of the questions singly or together. It is not his business to adjust his marking to suit his idea that the question was too hard or too easy, was outside, or only in part within, the syllabus. The question has been set and answered, and his task is to assess the relative merits of the answers.

He is not concerned with the candidate's teaching. He must neither penalise the candidate who he considers has been badly taught, nor give extra marks for evidences of what he holds is good teaching.

Where two methods are equally admissible, since they lead to the desired result, he must not favour one of them. It is not his business to judge that a candidate has been well or ill taught; he merely records the scores. He is not concerned with the orthodoxy or the beauty of the batsman's style; he merely counts up the runs. As a scorer it is of no import to him whether the batsman is a Jessop or a Hobbs, or, even, a Rabbit.

If the examiner is one of a team of examiners, then his query to himself at every little difficulty which arises must be, 'How would the others deal with this script?' If he is at all in doubt about any mark he scores he must not hesitate to send the script to the chief examiner for a revision of his scoring.

Further, the examiner becomes a machine. He decides that an answer falls into a category specified in the schedule of marking, and attaches it to the appropriate score. He must feel that his colleagues would concur in his decision. If he does not feel this certainty the script should be referred to the chief examiner.

Preliminaries. Objective marking implies a preliminary survey of the scripts to discover what the candidates have done. This survey is by sample. Shortly after the examination sample bundles of scripts are scrutinised by the examiners in the light of a preliminary suggested schedule of marking. The preliminary schedule compiled by the chief examiner suggests a distribution of marks. The preliminary survey is intended to find out whether this schedule

covers all types of answer, and whether the scores suggested are likely to produce the desired spread of candidates. During the preliminary survey the assistant examiners make careful notes of the work done, and mark the scripts tentatively. In regard to certain questions the frequency with which an answer occurs is noted.

Some days later the chief examiner presides over a conference of examiners, at which the suggested schedule is discussed piecemeal, amended, and approved. This conference is the most important part of the work. The conference finally determines the schedule of marks.

The marking. Assistant examiners mark the scripts in accordance with the schedule, from which no intentional departure is permitted. When an assistant examiner (i) finds a case outside the schedule ; (ii) would prefer for special reasons to depart from the schedule in a single case ; (iii) is in doubt about the intentions of the candidate, he marks the answer and initials his mark. He then puts a special sign on the script, which eventually goes specially to the chief examiner for his decision on the initialled mark.

These few words describe the bulk of the examining work. The assistant examiners mark thousands of scripts in a limited time, frequently during intervals of a permanent occupation. They work at high pressure at what is in reality a routine, somewhat mechanical, task.

The checking. All examiners make mistakes ; the good examiners in a small percentage of scripts

only. Some of the mistakes are purely clerical in the totalling or entering of marks ; others are lapses from strict adherence to the schedule, due to the high pressure at which they work. To a chief examiner these lapses are an interesting psychological study. There is no possible guarantee that a batch of scripts sent in by an assistant examiner will be free from error. A staff of checkers is employed to examine the marked scripts for possible clerical mistakes and for departures from the schedule which have not been initialled. In all cases where the checker finds *prima facie* evidence of an error in marking, the script is referred to the chief examiner, who alone makes alterations (where necessary) in the marks.

The clerical staff. The clerical staff takes the mark sheets of the several examiners for the separate papers, enters them on a total result sheet, adds the marks, and checks these operations. It makes any adjustments in the marks of one paper which are necessary to preserve the balance of the papers, using a correction or adjustment table prepared by the chief examiner. It analyses the total marks, and finds out the limit marks within which is the borderland. It then sorts out the scripts of the borderland candidates.

The borderland. The borderland includes x candidates who are just in the list of successful candidates on the first marking, and x candidates who are just as much out of the selected list. Probably the most important part of the work of a chief examiner is to consider the borderland cases and decide, chiefly

from consideration of the whole performance of the candidate, whether the first mark shall be altered. He pays particular attention to marks which have been initialled. In this connexion he is helped considerably by an intimate acquaintance with the character of the work of the assistant examiners, with their foibles and tendencies.

His review of the borderland completed, the necessary changes on the result sheet are made, and the order of merit list is completed in so far as the scoring of the marks is concerned.

Marking intelligence tests and Arithmetic attainments papers. In intelligence-type questions the answer is usually marked right or wrong, and the preliminary survey of the scripts is limited to a search for answers which can be regarded as correct. The actual marking of the scripts is a clerical task, largely a matter of eyesight ; it is certainly not an exercise in the interpretation of a candidate's words in relation to a schedule of marks. In fact, it is possible for a paper of intelligence-type questions in German to be marked by a person ignorant of German.

An Arithmetic attainments paper of sixteen to twenty short sums is usually marked 'right or wrong,' since only a mark or two is scored per question. An attainments paper of five, six, or seven longer sums requires a longer preliminary survey and a lengthy conference of examiners to determine the schedule of marks. No allowance of marks is usually made for method, since such an allowance would be examining teachers instead of children.

Two generally accepted decisions are important: (i) an incorrect step involves only the marks attached to that step; (ii) mental work, *i.e.* a solution without any working, is either right or wrong, and scores maximum marks or zero. The first decision means that if a candidate makes an error in the first step of a solution the examiner has to check all the subsequent working, and, if correct, score the marks attached to the steps involved.

Marking English questions. A paper of English questions requires a complete and exhaustive preliminary survey, since there are usually many possible solutions, all worthy of full marks, and other solutions worthy of some marks. The chief examiner is rarely able to suggest an *à priori* schedule of marks. All he can usually do is to suggest the allocation of marks to questions and parts of questions, and to provide a hint or so of the kind of solution probable.

The preliminary objective survey yields a large quantity of material which is reduced to some sort of order at the conference of examiners. The examiners must agree upon a schedule of marking which is sufficiently detailed to cover the cases discovered, sufficiently explicit to leave no room for disagreement between the markings of different examiners, and yet sufficiently brief for the examiners to remember during the process of marking, so that they are not required to refer constantly to the schedule. At the examiners' conference the question of the marks to be scored by certain solutions is frequently solved by the frequency with

which the solution occurs. For example, 90 per cent. of the scripts contained the word 'safety' as the solution to an English question; the majority of the examiners held the opinion that 'safety' was a good but not the best possible solution; the decision of the conference would necessarily be that 'safety' must score full marks because the candidates overwhelmingly considered it the best possible solution. This is an example of objective marking.

Frequently the Advisory Board will accept a suggested question in English, although the question seems to be rather difficult, because they rely upon the principle of objective marking to secure even justice to the candidates.

In practice the great difficulty in relation to English questions is to allocate the marks so that finally the average mark and the spread of the marks among the candidates are suitable. Should these prove in the outcome unsuitable, they must be adjusted mathematically. One of the guiding factors in the chief examiner's scrutiny of the borderland scripts is the circumstance that the paper of English questions functions correctly in the whole examination. He cannot accept the final marks on this paper without careful consideration of the incidence of the marks, say, between 70 and 80 per cent. of the maximum in relation to all the work.

It is an instruction to examiners to call attention specially to any script where the total mark disagrees with the general impression formed by the

examiner of the capacity of the candidate. In practice this instruction operates more frequently in the English questions paper than elsewhere, and the recorded opinion of the assistant examiner becomes valuable if the script concerned lies within the borderland.

Marking scholarship History and Geography. Objective marking for History and Geography for scholarships requires :

(i) an allocation of marks per question which will produce an average mark and spread of marks which are in harmony with the other papers ;

(ii) decisions of a slightly different character in relation to three types of questions :

(*a*) questions involving matters of fact,

(*b*) intelligence-type questions,

(*c*) essay-type questions.

In the case of matters-of-fact questions the marks depend upon the degree of precision required, *i.e.* upon the limits of error which are to be accepted. The preliminary survey alone decides just exactly how rigid the examiners should be. For example, the candidates are requested to locate a certain town on a map, say, Birmingham. The preliminary survey will discover the size of the rather large area in which the majority of the candidates who have a fairly correct idea of the location of the city place their dots. At the examiners' conference the area to be accepted is determined.

Answers to intelligence-type questions should be right or wrong, but it frequently happens that the words used to convey the idea which is probably right fail in their object. The preliminary survey will discover the common vocabulary used, and the conference will decide the issue. It may happen that the examiners will decide to accept a phrase which they would condemn in class use were they teaching the class.

In the case of young children competing for scholarships, essay-type questions are not easy to answer, and not easy to mark. They leave a great deal of scope to the examiner to examine the teacher and not the candidate. It would seem that the guiding principle to adopt is to mark for atmosphere, in Geography for the sense of relation in space, in History for an appreciation of the relation in time. The preliminary survey should yield an impression of the average way in which each question has been answered, with some notion of the best and worst answers. A question sometimes demands the statement of certain facts. The preliminary survey will settle just which facts have been mentioned and the frequency of their occurrence. The conference will discuss these observed results and decide on the marking. In this way the apparently hard question loses its difficulty.

Marking other papers. The marking of essays and Arithmetic problems is discussed in the chapters which follow. The marking of certificate scripts presents individual difficulties in view of the twenty different subjects, and in view of the fact that in

each subject there are two or three levels of performance which matter. For example, as far as the result of the examination is concerned, if 70 per cent. scores distinction it does not really matter whether a candidate scores 80 per cent. or 90 per cent. The borderland, 68 per cent. to 72 per cent., however, matters greatly.

The general principles and practice of objective marking have been illustrated in this chapter, and their application to the marking of French, Chemistry, etc., is of particular rather than general interest.

The chief examiner for a certificate subject has to divide the scripts thus :

(*a*) Certain distinctions.
(*b*) Borderland distinctions.
(*c*) Certain credits.
(*d*) Borderland credits.
(*e*) Certain passes.
(*f*) Borderland passes.
(*g*) Failures.

He is not concerned with the balance of his paper among the other subjects. In that regard he merely obeys instructions.

It is probable that certificate marking would gain in precision if literal marks were used instead of numeral marks. For example, the seven letters (*a*) to (*g*) *supra* might be appended to each answer on a script, and the ultimate fate of the script might be recorded by one of these letters. There is no fundamental necessity that certificate answers should be

marked numerically; in fact, an examiner would face a hard task to try and justify the difference between 14 and 15 and 16 marks, say, out of a possible maximum of 30. All these are 'borderland credit' marks—the script requires a second opinion at least.

VIII. ESSAYS

An essay paper. The fundamental consideration is that an essay paper shall provide the candidates with a stimulus towards the production of a passage of connected prose. The passage may be a piece of quite free composition in response to such a question as, 'Write on any subject you please'; or it may be the reproduction of a passage read. Sometimes the stimulus is provided by a picture or set of pictures which give rise to a story. More often, however, the candidate is faced with a choice between two or more subjects, one of which is rather matter-of-fact, while another is more imaginative.

The choice allowed in an essay paper is a compromise between two notions—one that all the candidates shall be stimulated by at least one of the subjects, the other that the candidates shall not be embarrassed by a wealth of options. The number of options depends largely upon the age of the candidates.

Freshness. The essential factor in an essay subject is freshness, and the provision of the necessary stimulus which has an element of surprise for the candidates is one of the ways in which the chief examiner and the teachers of the schools enter into

competition. Teachers use dozens of subjects in their routine work, and probably use the subjects set in the past by the chief examiner, and, further, attempt to anticipate the subjects which the chief examiner is likely to set in the future. The chief examiner, in opposition, must attempt to provide a stimulus which has not been anticipated, so that the candidates may experience a feeling of novelty when they read the paper.

The members of the Advisory Board, being adults, sometimes feel that the subjects proposed for their adoption are rather dull, yet experience shows that dullness is a relative term, and that subjects which were thought to be dull have been successful examination subjects, *e.g.* 'The people who live in our street.'

The subject chosen for a scholarship paper does not matter greatly, provided it stimulates the candidates to write freely. If a subject threatens to be difficult because it is somewhat outside the experience of the candidates, the difficulty may be overcome by providing for previous perusal a passage on the subject in order to give the candidates ideas which they may or may not use. A passage so provided is not meant to suggest an exercise in paraphrase ; it is not intended that the candidates shall memorise the ideas and vocabulary and reproduce as many of the ideas as they can. It is simply a form of stimulus which excites the candidates to write.

The general nature of an essay test. Analyses of scholarship examination results for examinations

for different areas, at different ages, and for different sexes, indicate that the essay marks correlate closely with the marks scored for the whole examination. The essay paper acts as a kind of controlling governor on the whole examination and steadies the results. It counteracts the special tendencies of Arithmetic attainment and English questions papers, and gives the capable candidate of general capacity a chance to score readily.

Boys and girls. Current tradition avers that there is a subtle distinction between the essays of boys and girls on the same subject, and that girls score more freely than boys. Experience provides many indications which suggest that this tradition is wrong.

Skilled examiners experienced in reading essays provided by several examinations of the same type have been asked to distinguish the sex of the writer from reproductions of sample essays on the same subject. They were as likely to be right as wrong, and there was no consistency in their verdicts.

The detailed analyses mentioned in the preceding paragraph provided no invariable indication that boys scored less freely than girls. The analyses, however, suggested that a difference between boys and girls arose in a small measure as a reflex of the personality of the examiner. On the same paper at the same examination boys scored more freely than girls when the scripts were marked by examiners *A* and *M*, and not when marked by examiners *B* and *N*. Examiners *A* and *M* were not of the same sex, neither were *B* and *N*. Some men examiners

seemed to be inclined to favour the boys, other men the girls, and so on.

The probable solution of the sex question appears to be that the tradition has been manufactured by the examiners of the past, and that some examiners of the present, consciously or unconsciously, are biased towards or against the girls, and that it is really the examiner's bias which is responsible when differences occur. In any event, the bias of an examiner is a definite factor which has to be taken into account.

Marking essay scripts. The chief decision in relation to essay scripts is that no candidate shall be considered among the certain winners or in the borderland until his essay has been read at least twice independently. This decision is inevitable, because experience shows that there is always a small proportion of essays on which two examiners will disagree somewhat greatly. There was always the chance that a successful candidate might owe his success to the luck that his essay was read by examiner *A* and not by examiner *B*. The second reading removes this element of chance.

When the two scores vary greatly the essay is read by a third examiner, and in the borderland the history of the final mark scored for the essay is one of the items which engage the chief examiner's attention. The wisdom of these decisions is demonstrated by the close correlation between the essay mark and the total examination mark.

A schedule of marking. It is highly probable that the tradition that essays are difficult to mark

consistently arose because examiners were not in sufficiently close agreement upon a schedule of marking and upon what qualities in an essay were to count. The old examiner was in the habit of bunching his essay marks close to the average, with a very narrow spread, and it is on record that an examiner owed his reputation as a skilful reader of essay scripts to the fact that most of his scripts scored somewhere about half marks, and a few definitely good and a few definitely bad scored other marks.

The modern examiner requires a wide spread of the marks with a view to a reasoned estimate of the differences between candidates more or less of average ability, because these differences are important for borderland cases.

The old examiner marked largely by impression of the essay as a whole. One method was to read the scripts once and mark them α, β, or γ; then to read them again and mark with + or − in addition to the α, etc. Sometimes + + or − − were used. In this way the examiner got a scale of fifteen divisions to which marks were attached, *i.e.* his steps of difference jumped by 7 per cent. This method had one drawback, that the examiner had to read batches of scripts, and so completed perhaps one batch of five hundred scripts before he started on the next batch, with the result that he was never in a position to revise his marking in order to spread his marks widely. His spread of the marks was mathematically adjusted by the chief examiner.

A method in use by the modern examiner is to

mark by points. His working maximum is 50, a mark which is reserved for those few scripts which are too good to require analysis, zero being correspondingly used for those few scripts which are too poor to bear inspection. He marks 7 points, to each of which he scores 1, 3, 5, or 7, which means that his steps of difference jump by 4 per cent. His 7 points are :

(*a*) Vocabulary	Technical skill.
(*b*) Accuracy	
(*c*) Craftsmanship	
(*d*) Consistency	Ideas.
(*e*) Completeness	
(*f*) Quantity of Ideas	
(*g*) Quality of Ideas	

(*a*) Vocabulary refers to quality and quantity of words used.

(*b*) Accuracy deals with technical errors in spelling, grammar, etc.

(*c*) Craftsmanship refers to the quality of the phrases and sentences.

(*d*) Consistency refers to irrelevancies and confusion of statement.

(*e*) Completeness deals with the unity of the essay, its beginning and its conclusion.

(*a*) to (*e*) Five points : marked irrespective of the subject set. This means that an essay completely off the subject set would be marked to a possible maximum of 35.

(*f*) and (*g*) Two paints which explain themselves. The relevance of the essay to the subject set solely

affects the mark for point (*g*), and marks are only scored for (*f*) for the number of ideas relevant to the subject set.

When marking, examiners may work to a foot-rule. Of every 200 scripts or so they select 14 at random, and count up the total number of times the marks 1, 3, 5, and 7 have been scored. If there are x 1's, there should be x 7's; $(49-x)$ 3's and $(49-x)$ 5's. The values of x for different samples of 14 scripts should fluctuate about 19.

The preliminary survey, etc. Each examiner is provided with a sample selection of scripts, which he reads, and for which he scores marks provisionally. He chooses 2 essays for each of the 7 points, one worth 3 marks and the other worth 5 marks, and sends the 14 essays to the chief examiner.

The chief examiner compares the essays from the different examiners with sample essays which he has provisionally marked in the same manner as his colleagues.

At the conference of essay examiners there is a general talk upon the scripts read with a view to standardising the idea of an average performance, and the 14 essays are handed back to their respective examiners with comments where necessary. A free and extensive discussion of all matters noted by the examiners is desirable.

After the conference, examiners mark a batch of 50 essays, which they send to the chief examiner post-haste. He reads these, compares his marks with the examiner's, and returns them with any comments which may be desirable.

Revising examiners. Revising examiners take part in the preliminary survey and conference. They receive scripts upon which the first reader has made no mark, and work to a foot-rule which is adjusted in relation to the minimum mark which qualifies an essay to be read a second time. They are not informed of the marks scored on the first reading.

When the revising examiner's marks are compared with those scored at the first reading it is usually found that the incidence of fluctuation between two scripts is the same, *i.e.* both examiners agree that script A is better than script B, and so on. The degree by which A is judged better than B varies usually by not more than a mark or two, so that the number of scripts sent for a third independent reading is small.

The 7-point system. The system described in the preceding paragraphs succeeds in its aim. It keeps a team of examiners consistently together; it causes the average candidate to score approximately half marks; it spreads the marks in the fashion which the modern examiner desires. Finally, its results correlate closely with total marks for the whole examination.

Naturally enough, examiners find such a scheme novel and, at first, difficult to work. But it forces them to do what essay examiners seem to dislike—to score 7's for points when the essay is well above the average in merit, and to score as many 1's as 7's. The system has been tried out with four different teams of examiners for different examinations, and the results each time were as recorded above.

The scheme has an additional advantage in that it yields upon analysis valuable information about the habits and idiosyncrasies of examiners—information which is simply invaluable to the chief examiner when he scrutinises the borderland.

No claim is put forward that the scheme is suited for marking essays other than those written by young children in scholarship examinations. In these examinations the competition is so keen that every mark counts, and so every mark must withstand scrutiny, not only when borderland cases are under consideration, but at any time, years hence, should the occasion arise. Many teachers and examiners will feel that the scheme leaves out certain somewhat indefinite but desirable qualities which should count in a good essay. To meet such a feeling, and to prevent the possibility of such excellent but elusive qualities failing to meet their due reward, examiners are instructed (i) to send for special consideration to the chief examiner any script which, in their opinion, deserves marks different from those obtained under the scheme; (ii) to refrain from analysing those essays which they deem worthy of 43 to 49 marks or unworthy of more than 6 marks.

The marking of similar subjects. As has been previously noted, some authorities hold examinations for special scholarships in which Needlework, Drawing and Woodwork are examined. These subjects are as difficult to examine as essays, provided it is required to place the candidate's work in a relative order of merit, with a wide spread of the

marks. All these subjects are difficult, because the judgment of two examiners is not likely to agree exactly, and may vary in odd cases considerably.

For scholarship purposes the methods outlined for essays should be applied to Needlework, etc.

In History and Geography papers, were sufficient time allowed and sufficient money available to pay extra fees, it would be desirable to apply these methods to the marking of all essay-type questions. Since neither time nor money are available, the makeshift scheme already suggested must be adopted, and the chance of serious error minimised by the judicious use of matter-of-fact and intelligence-type questions.

IX. ARITHMETIC PROBLEMS

The aim of the paper. The paper in Arithmetic problems is designed to give a chance to the really capable child of scoring heavily. Recent papers have set five problems to be solved in forty minutes, and, as usual, some boys and also some girls have scored full marks because they have solved all the problems.

The problems set are difficult, but they have not yet been so difficult as to find the upper limit of the ability of the brightest English children of 10+ years.

It is an axiom that candidates who score full marks are not really tested by the paper on which they score the possible maximum, and, therefore, there are some children at every examination who are not really completely tested, and there is no indication of how much better they could do.

The type of the paper. Although the paper is labelled Arithmetic because quantities are introduced, it is first of all a test in English ; the candidates are required to pass a comprehension test. Secondly, they are required to think quantitatively, selecting the data they need, and disregarding anything else. Thirdly, they must state their solution

and give some evidence of the way in which they reached the answer.

The type of question. For children of 10+ a problem should

(i) deal with small quantities,
(ii) require computations which can usually be performed mentally,
(iii) involve a principle,
(iv) be capable of an immediate solution by direct reasoning, and also of an indirect solution obtained by teachable methods.

The examples which follow will indicate how these factors may be combined.

What problems test. Primarily these problems are intended to test insight or vision, the ability to go to the root of the matter and elicit a clear issue from a number of words. Adults find these problems more difficult than capable children, because their insight is dulled by their greater acquaintance with the world and its possibilities. While an adult hesitates where to begin, the bright child goes quickly to the essential data. The human quality, here dubbed 'vision,' is not teachable. It may be an educable ability, but it is a power which the teacher continually meets. Some children 'see things at once'; others see them only when presented with the assistance of routine explanations, *i.e.* in relation to past experience; others fail to see them at all.

Secondly, the problem is a test of mental effort

or concentration. The scripts indicate clearly for every problem the stages of concentration of which the candidates are capable.

Thirdly, the problem is often a test of the power to state an argument, for the candidates are sometimes specifically directed to give reasons for their answers, or to demonstrate why a certain solution is correct. This ability, which develops late in children, is an indication of the capacity which the new examiner implies when he states that the mental age of a child of 10 years is 15 years.

Marking problems. Most problems yield four or five solutions, of which one is correct, and the others indicate types of thinking which betray the power of concentration of the child. The frequency of these four or five solutions is more or less constant for all problems, and it is rarely that the answers cannot be classified in this way. The preliminary survey is intended to detect these 'type' solutions. For example:

Question. Two shopkeepers bought 5 cwt. of sugar, each paying half the cost. Instead of dividing the sugar equally, one took 3 cwt. of sugar and gave the other £1 3s. 4d. to make things fair. What was the price of sugar per lb. ?

The solution depends on the weight of sugar which the candidate considered was equivalent to £1 3s. 4d. The weights which the candidates actually set down were, in cwt., $\frac{1}{2}$, 1, 2, 3, any other, none.

Actually in percentages the proportions in which these weights were set down were :

	½	1	2	3	Any other.	None.
Boys -	23	21	13	9	18	16
Girls -	17	19	16	8	25	15

The examiners' conference decided what mark shall be attached to each type solution, and the schedule of marking for the above question contained 6 values for the types specified above, provided the subsequent change from x cwt. = £1 3s. 4d. to a value in pence per lb. was correctly computed ; and a statement of the deductions to be made when this computation was incorrectly made. The examiners had, therefore, at least 12 mark values to determine at the conference.

A second example :

Question. A man walks a certain distance at a speed of 3 miles an hour, and cycles back at 9 miles an hour. He takes 8 hours for the double journey. What fraction of the whole time does he spend in cycling ? How far does he walk ?

Two answers are required, and the preliminary survey showed :

(*a*) that to the first question the answers $\frac{1}{4}$, $\frac{1}{3}$, $\frac{3}{4}$ had some merit, and that 37 per cent. had other answers than these, and 22 per cent. reached no answer at all.

(*b*) that the second answer among the 41 per cent. who reached a reasonable first answer was

consistent with the first in 23 per cent. of the scripts and inconsistent in 18 per cent.

The schedule of marking showed the marks to be attached to the first answers, *i.e.* 3-mark values, and the additional marks for a second answer consistent with the first; and, in some cases, for a correct second answer inconsistent with the first answer, or, in other cases, for certain second answers which were incorrect and inconsistent with a correct first answer.

These two examples indicate how complicated a schedule of marking becomes, and they show further :

(i) that the marking of Arithmetic scripts is not inherently easy because the subject is Arithmetic.

(ii) that consistent marking of Arithmetic requires a careful preliminary investigation and a strict adherence to the schedule.

(iii) that it is within the power of the examiners to attach such marks to these ' type ' solutions as will produce almost any distribution of marks they desire.

(iv) that, in fact, the marking of Arithmetic problems is as difficult as the marking of essays, and leaves just as much scope for the personality or bias of an examiner to produce unwanted inconsistencies in marking unless the examiners agree on a schedule and honour their agreement to the letter.

(v) that checking the marking of these papers in Arithmetic is a tedious yet very necessary

part of the total work attached to an examination.

Observations on specimen questions. The following notes on questions which have been set recently are apposite :

Question. A number of children were invited to a Christmas party. Enough oranges were bought for each invited guest to have two. Some children were absent, so that a quarter of those present had a third orange. What fraction of the children who were invited were present ?

The way in which an early slip frequently adds to the difficulty of a sum is illustrated here. A few candidates misread ' a third orange ' as ' a third of an orange.' Some of these reached a solution correctly based upon this misreading, no small achievement in an examination room.

The common answer to this problem was $\frac{3}{4}$, an answer which lacks justification ; less frequently the answer was given as $\frac{4}{5}$, based upon the interpretation that one orange each was provided, and that a quarter of those present had a second orange.

The amazing ability of the best candidates is evidenced by the following actual solutions :

(*a*) Four children 9 oranges instead of 8.

Ans $\frac{8}{9}$.

(*b*) Child had $2\frac{1}{4}$ oranges instead of 2.

Ans. $\frac{8}{9}$.

These solutions exhibit the insight, or vision, which the problem was set to detect. They are such solutions as an average class of children aged 10+ would not readily credit as correct. The average child would require one of the 'teachable' methods of solution such as a 'diagram' method for the brighter children, or a 'hypothesis' method, such as :

Suppose 16 children were present.
4 had 3 and 12 had 2 oranges—total, 36 oranges.
36 oranges were provided for 18 children.
Ans. $\frac{16}{18}$ or $\frac{8}{9}$.

The difference between the vision of the brilliant child and the painstaking care of the well-trained child to whom taught methods alone appeal is strikingly illustrated in the following :

Question. A boy does $\frac{2}{3}$ as much work per hour as a man. There are two gangs, one with 2 men and 3 boys, and the other with 3 men and 2 boys. Which gang will complete a job sooner ? Show how you got your answer.

It is difficult to improve upon the solution :

There was an odd man and an odd boy.
Ans. The gang with more men.

The teachable solution which requires the computation that one gang is equivalent to 4 men, while the other is equivalent to $4\frac{1}{3}$ men, involves a method which children of 10+ find difficult. This was clearly indicated on the scripts by the comparative

frequency with which the computed values 4 and $4\frac{1}{3}$ led to the wrong answer.

Question. A man is paid 5s. a hundred for making toys, and earns £5 a week. He works a machine 8 hours a day five days a week. If he worked the machine twice as fast and only worked 6 hours a day, how much more would he earn in a week ?

The solution :

6 'double' hours = 12.
12 is 4 more than 8.
Increase is by a half.
Ans. £2 10s. 0d.

means the mental elimination of data which are required for the following, more common, 'step-by-step' solution :

At first	(i) per week,	$20 \times 100 = 2000$
	(ii) per day,	400
	(iii) per hour,	50
Later	(iv) per hour,	100
	(v) per day,	600
	(vi) per week,	3000
	(vii) earnings, -	£7 10s. 0d.
	(viii) increase, -	£2 10s. 0d.

Question. A cinema which has 300 seats charges 5d. for adults and 3d. for children. When the cinema is full, does it pay better to have $\frac{1}{3}$ or $\frac{1}{2}$ of the seats occupied by children, and what is the difference in the takings ?

The child with sufficient insight to select the data which mattered gave the solution :

A third by 50 times 2d., 8s. 4d.

Less capable candidates tried to work out all the details, with the result that some of them lost their way in an unaccustomed mass of figures, and gave up the task, while others reached incorrect solutions.

Freshness. For many reasons problems papers in Arithmetic must present an element of surprise. The questions must be fresh to the candidates. Some teachers use these problems as a kind of 'type' sum, and as a basis upon which to spot likely problems in future papers. This fact forces the setter of the problems to present his problems in so novel a manner, *i.e.* a fashion so far removed from the conventional style, that the superficial and thoughtless adult—sometimes a parent, sometimes a journalist in search of 'copy'—considers the problem as a freak question. The test of what may be called the legality of the problem is to try and solve it by algebraic, diagrammatic, or general methods with letters, lengths, or areas, in place of the given quantities.

The fact that many of these problems yield an elegant algebraic solution has suggested to many teachers that it is desirable to introduce algebraic methods into the teaching of children of 10+ years. Such teachers should stop and think over what they are attempting. Theoretically it is scarcely to be expected that a child of such tender years is sufficiently well developed mentally to be able to solve

problems algebraically, and to force a child to use 'tools' instead of inductive insight cramps the child.

Practically, the extension of such a practice to all the schools would force the setters of problems to forsake such material in future papers, and the child would become 'over-trained' and 'stale' for no purpose.

Experience shows that children of 10+ who attempt to use algebraic methods in solving the problems almost invariably fail to score any marks at all.

An adult finds these problems difficult because he has to chose his method of trying to reach a solution. A child is not penalised by the hesitation between two or three possible methods of attempting a solution ; and the teacher who tries to put the child in the position of an adult is spoiling the child's chances.

Intelligence-type questions. Intelligence-type questions and solutions by mental arithmetic are not admissible in connexion with Arithmetic problems. The essence of the matter is to get the candidate not only to solve a problem, but to give also some indication of the mental process by which he reached the solution. It is not anticipated that the statement of the solution will be deductively beyond reproach ; it is not expected that the solution shall be elaborately presented in a mass of verbiage ; but it is expected that the steps leading to the solution will be indicated in a manner which the examiner can readily follow. The teacher who sometimes sets a

problem and also gives the answer, and then trains his pupils in methods of demonstrating why the answer is correct, prepares the way for deductive statements such as the most brilliant candidates present on their scripts.

An Arithmetic problem paper is useful to the examiner, because it lifts the most capable candidate well above the average, and because it tends to differentiate on the basis of the character of the solutions submitted between degrees of partially successful reasoning. It frequently happens that a candidate has some idea of the solution, yet first misses the correct notion. His answer has more merit than an answer which is muddled or only proceeds a step or two towards the solution or shows a fundamental misconception.

In practice it happens almost invariably that candidates who score well on this paper score well throughout the examination. There is a close correlation between high marks on this paper and scholarship success. The correlation, however, is not perfect, because some candidates score well enough on the other papers to reach scholarship level, although they score but a few marks on Arithmetic problems.

It is considered that so long as the rest of the examination gives scope for these children of more ' pedestrian ' capacity this paper is justified by the ' flights ' of intellectual capacity which it produces. In the borderland marked ability in this paper is rated rather highly, so that a candidate who scores really high marks in Arithmetic problems, and only

succeeds in scraping into the borderland, is held to be a better candidate than one who scores a dead level of performance on the other three papers with a small score in Arithmetic problems.

A curious phenomenon is beginning to make its appearance. Some candidates are scoring more easily on Arithmetic problems than on the Arithmetic attainments paper. This phenomenon is one of the reasons for the opinion that it should be possible to set papers in Arithmetic on which boys and girls can score with equal facility.

X. PREPARING THE LIST OF AWARDS

Compromise. When the marking has been finished and the clerical work of totalling and checking, etc., has been completed, the work of the examiner is in reality concluded. The award list is a compromise. Various authorities take other factors into account in different degrees. Some authorities throw the onus of making the compromise upon the chief examiner, since he alone is acquainted with the complete performance of any candidate.

The extraneous factors which call for consideration are :

(i) a special report by teacher or inspector upon an exceptional child.

(ii) lists, in order of merit, compiled by the head teachers of the schools which submit the candidates.

(Both (i) and (ii) are made before the examination takes place, and are not consulted until the ' crude ' list of awards has been compiled.)

(iii) specific excellence indicated by high marks and the quality of the work in either English or Arithmetic, coupled with mediocre ability in the other subject.

(iv) a rare ability manifested solely by very high marks on intelligence tests alone.

(v) an allowance for age differences.

Certain successes. The 'crude' award list can readily be divided into 'certainties' and 'possibles.' The certainties are those candidates who reach scholarship level (*a*) on the whole examination; (*b*) upon English separately; and (*c*) upon Arithmetic separately. The 'possibilities' are those candidates whose all-round average brings them up to scholarship level, and those who score much better in one subject than another. The compromise affects only the 'possibilities.'

Exceptional children. Experience shows in general that the exceptional children reach scholarship level, and that the special individual recommendation was not necessary. It suffices, as a rule, for the head teacher to place such children first in his order of merit list.

Head teacher's lists. The list in order of merit compiled by the head teacher, usually in consultation with his colleagues, is extremely useful in the case of the possibilities. An analysis of head teacher's lists and examination lists in order of merit indicates that the two lists agree quite closely concerning the candidates who are likely to gain awards. As a rule examiner and head teacher agree upon the relative merits of the candidates.

In practice the examination really decides that:

(*a*) No. 1 on some head teacher's list is not of scholarship level.

(*b*) No. 1, 2, and 3, say, on other lists are all possibilities.

(*c*) No. 1, 2, 3, and 4 (perchance) on yet other lists are certainties, and 5 and 6 are possibilities.

The compromise works out that the head teacher's opinion as recorded in his list will turn some possibilities into certainties, and will bring a No. 1 into the award list when otherwise he would have been just below it. For example, the examination says No. 2 and 3 just qualify, and No. 1 just fails to qualify; then the head teacher's opinion should suffice to bring No. 1 into the list of awards.

Specific excellence. In general there are few cases —not more than 2 per cent.—where a candidate who scores really well on one subject or on intelligence tests fails to secure an award by a really poor performance in the other subject. For example, when the scholarship level is 150 marks out of 200, it is rare for a candidate who scores 90 in one subject to fail to score 60 in the other.

Yet there are cases of candidates who score full marks on one paper and fail to get into the ' crude ' award list. There is a chance that such candidates have not been completely tested, and that the full marks indicate exceptional ability. Such cases are considered on their individual merits as shown on the scripts. Full marks on an essay or on Arithmetic problems would in most cases justify a compromise and the inclusion of the candidate in the award list.

Age allowances. Age allowances are usually automatic. They may take many forms, some-

times a bare addition or subtraction from the candidate's score; at other times the change may be a 'percentage' of the score instead of a fixed number. For example, in one scheme a candidate might receive 6 more marks; in another scheme he might get 6 per cent. more marks, thus:

$$150 + 6 = 156;$$

or 150 + 6 per cent. of 150 = 159.

The decision on age allowances rests with the authority. In default of a careful and complete investigation into the subject of age allowances—an investigation which is overdue—each authority must find a scheme which yields in practice results which seem satisfactory. The crude criterion seems to be that the awards when the candidates number thousands should be proportional to the number of days in the months, *i.e.*, if 14 children born in February receive awards, 15 children born in April, June, etc., should receive awards. There is no evidence, however, that this crude criterion has any merits.

The final list. When the compromise has been effected the list is endorsed by the Advisory Board, which may be required to make the final decision in regard to a few candidates. It is then submitted to the authority.

XI. SCHOLARSHIP EXAMINATIONS AND THE ELEMENTARY SCHOOLS

Preliminary tests. When the number of children within the age group which is to be submitted to a scholarship test is very large, it is desirable to make a preliminary selection so that the weaker children may be eliminated. This course is desirable on the ground of economy, and also because examining too many candidates interferes with the success of the examination by adding to the concentration of effort which is required to get through the work in a short time.

Some authorities base a preliminary selection upon the position of the child in school. For example, there are children of 10+ in Standard VI. and in the three standards below, the term standard being used to indicate approximately the relative position of the children. By the decision that all children of 10+ shall be examined, provided they are at least in Standard IV., some 25 per cent. of the children in the age group are regularly eliminated.

The preliminary weeding may take the form of an examination with similar, but easier, papers to those of the main examination. An approximate

fraction of the children who take the preliminary examination is eliminated by taking the list in order of merit at the preliminary examination for each school and discarding those at the bottom. At such a preliminary examination the children compete only against children in their own school; at the later examination the competition is much wider and keener. At the preliminary less than half the children are successful; at the final possibly not more than 15 per cent. gain awards.

Other authorities rely on mental tests to determine which candidates shall be eliminated, and they use the results of these tests as a rough check upon the 'crude' award list, to ensure that no really able child is missed.

The calibre of the successful candidates. Consider the conditions which obtain under an authority which has awarded scholarships for many years as a result of the same sort of examination. Assume that the schools which regularly submit at least six boys and six girls to the examination have maintained their conditions without any radical change of organisation during the period under review. Approximately the candidates are time after time from the same schools, and are submitted to the same sort of test. Assume, further, that the examination is objectively conducted with the sole aim of discovering the relative ability of the candidates.

Now take the lists of awards for the period, and tabulate the successes per school for boys and girls separately, and see what happens. One investiga-

tion of this sort has been made, with the following results :

(*a*) In most cases the awards for boys and girls were on the average the same.

(*b*) In most cases the awards during the first half of the period were the same as during the second half.

(*c*) Where the awards were not the same for boys and girls the schools concerned were nearly all located in one section of the authority's area in a compact group.

(*d*) The awards usually went to districts ; from the set of schools with many awards regularly there was a steady decline to the district containing the schools with only an occasional award.

The results led to the hypothesis that on the average and in the long run the awards for boys and girls from the same school should be the same.

This hypothesis was checked by a special consideration of mixed schools in comparison with schools which have separate boys and girls' departments. There was no evidence that the differences in teaching staffs affected the awards, and the hypothesis was confirmed.

The practical conclusion is this, that the awards depend largely upon the calibre of the children, and only in a small measure on the teaching. A school in a certain neighbourhood will gain many awards consistently. If the awards for boys and girls do not agree, then there is some special reason for the

divergence. Some schools can never expect to secure an award except an occasional sporadic success with a child who is out of tune with the average calibre of the school.

This hypothesis suggests that the successes at a scholarship examination provide no evidence on which to base a criticism of the conduct of a school, for it would first have to be determined that a school *should* secure x awards on the average, and there is no means of reaching such a decision. Some schools probably gain p awards when they possibly should gain $(p+q)$ awards, and this fact means that other schools gain m awards when they should gain $(m-n)$ awards.

A second practical consequence of this hypothesis is that so long as the scholarships are open and general—*i.e.* not limited either in the school from which the candidate comes or to which he goes—the only safe method of award yet devised is the competitive examination objectively conducted.

Effect of the examination in the schools. The examiner warns teachers repeatedly that nothing in connexion with the examination is so settled and stereotyped as to justify special modifications in the curriculum to meet the assumed examination requirements.

An objective examination follows the schools in a vague and general way. The papers must be fresh, and to some degree novel. An idea which will develop into a useful examination question might be quite unfitted for school use. For example, the examination use of a quotation from Hood, 'I

know a pleasant land where omelettes grow on trees,' etc., does not justify the appearance in the classroom of a list of words abstracted from the comparatively rarely used culinary terms, such as ragout.

Pressure on the children. It is frequently lodged as a complaint against scholarship examinations that they have a pernicious effect throughout the elementary schools by forcing the children. It is assumed that teachers whose pupils gain one award per annum are going to sacrifice the best interests of some 40 other children to secure a second award for their school. It is alleged that pressure is put upon the head teachers of junior schools to promote able children too rapidly.

The examiner is not directly concerned with the truth or otherwise of these more or less specific complaints, but he is intimately concerned with the misconception upon which they are based.

The pace is set by the teachers. They lead; the examiners follow. The examiner is not interested in what the children should do; his main concern is to find the limits of what they can do, on the average, fairly well, and to set tests within those limits which will yield the examination results which he needs.

If the examiner has to say definitely who are the top 20 or the top 500 out of 3500 candidates, his methods are controlled by that fact. If he has only to determine the top 1200 out of 3500 his methods would be different. If the children suddenly failed to manifest the average ability to which he is accustomed, then the examination would be modified.

The examination is a very sensitive instrument which automatically changes with variations in the area under review. Every question comes up, after the examination results are known, for subsequent review. Has it been a good question ? is always asked, and the criterion is—Has the question helped or hindered the relative placing of the candidates ? No one can judge the quality of a question without access to the results.

A question is endorsed as *suitable* by the Advisory Board *before* the examination. It is labelled good, bad, or indifferent by the examiner *after* the award list has been compiled. The chief examiner notes the cases when a question has turned out badly in his reports to the Advisory Board.

Preparation. It is a tradition that examinations must be prepared for. Hence the crammer thrives. Most authorities expressly forbid special preparation or cramming for scholarship examinations. They say that the children should take the examination in their stride, and they conduct an examination in Arithmetic and English without prescribed syllabuses to gain that end.

Misguided parents and over-conscientious teachers, however, believe that a little bit of extra coaching, an hour or two of special preparation, may result in an individual child's success, and so the teacher sometimes defines special preparation in a specious manner.

Practical experience indicates that *cramming does not pay.* The scripts of any scholarship examination provide the evidence every time. It may be

suggested that the hypothesis that the calibre of the child determines his success shows clearly that cramming cannot pay, for what amount of cramming will add a fraction to the inherent mental power of the child? Teaching is an interaction between teacher and pupil, and examining is a reaction of the pupil to a stimulus; the examiner wants the reaction from the pupil alone, and the crammed child is choked, so that the reaction is slow and laboured. Cramming will not give powers of insight and vision to the child.

In some schools a bright child is never allowed to forget that some day he has to sit for a scholarship examination, and that for the honour of the school he must keep on trying *to do better than he can*. Some authorities have Honour Boards where the names of scholarship winners are inscribed; others give the school a special holiday to celebrate the award of a scholarship to a pupil. Both practices are forms of special preparation or cramming.

An examiner's plea. To the child a scholarship examination should be a routine incident in school life, an incident which by its very novelty becomes a joyous adventure. Should teachers cause the child to dread the examination day? The papers are a stimulus to endeavour, should the dread of making a mistake prevent the free play of the child's intelligence?

The examiner has faith that all his work tends only to the discovery of talent, and is nothing else than a quest for capacity.

Let the teacher have faith in the product of his

year-long labours, so that the native capacity of the child, fostered and guided by well-thought-out plans, which began to operate when the child first toddled to school, may respond freely to the stimulus, the exceptional stimulus, of this one day. Let the teachers forget or ignore the examination as much as possible.

The measure of an examiner's success is the degree to which the special influence of the teacher of the candidates is eliminated. Can teachers co-operate to this end ? An examiner yields to none in his appreciation of the extraordinary capacity which he encounters in the course of his quest, and the fact that he can discover such remarkable ability gives him faith in his methods. Can teachers have equal faith in their methods ?

APPENDIX.

SPECIMEN SCHOLARSHIP EXAMINATION PAPERS WITH SOME COMMENTS.

ARITHMETIC.

Arithmetic Attainments Papers.

Specimen A. (Time 45 minutes.)

1. Divide 2133549 by 207.
2. Add $\frac{1}{13}$ of £6 17s. 0½d. to $\frac{1}{17}$ of £35 5s. 6d.
3. Find the values, in lowest terms, of :
   (*a*) $\frac{3}{4} - \frac{1}{3}$. (*b*) $1\frac{1}{4} - \frac{5}{6} + \frac{7}{12}$. (*c*) $\frac{\text{3s. 4d.}}{\text{8s. 8d.}}$.
4. Find the values of :
   (*a*) $40 \times 0{\cdot}25 \times 3{\cdot}4$. (*b*) $213{\cdot}3549 \div 20{\cdot}7$.
   (*c*) $3{\cdot}14 \times 12{\cdot}56$ (correct to the nearest whole number).
5. Find the value of :
   (*a*) 7½ per cent. of £125.
   (*b*) 13 per cent. of 3½ tons (answer in cwts.).
6. Find the cost of :
   (*a*) 11 oz. tea at 2s. 4d. a lb.
   (*b*) 1 lb. 3½ oz. of cheese at 1s. 4d. a lb.
   (*c*) 9 score eggs at 2s. 3d. a dozen.
   (*d*) 3 gross pencils at 4½d. a dozen.
   (*e*) 5½ yards of cloth at 2s. 11d. a yard.
   (*f*) 7 gallons of stain at 8s. 3d. a gallon.

Total cost is not required.

7. How many feet of planking 11 inches wide will make a table top 3 ft. 8 in. × 4 ft. 6 in. ?

Comments.

This paper was set at an examination in which candidates also worked an Intelligence Test in numbers and an Arithmetic Problem paper. The average age was 11 years.

SPECIMEN B. (Time 40 minutes.)

1. Add $\frac{1}{7}$ of £4 19s. $5\frac{1}{2}$d. to $\frac{1}{13}$ of £6 11s. 1d.
2. Divide 22,363 by 107.
3. Find the values, in lowest terms, of :
   (*a*) $\frac{3}{4}+\frac{2}{3}-\frac{5}{6}$. (*b*) $\frac{1}{4}\times\frac{3}{7}\times\frac{14}{15}$. (*c*) $\frac{\text{2s. 8d.}}{\text{8s. 8d.}}$.
4. Find the values in decimals of :
   (*a*) $40\times0{\cdot}5\times3{\cdot}3$. (*b*) $2{\cdot}4\div1{\cdot}25$. (*c*) $4{\cdot}04\times3{\cdot}75$.
5. Find the values of :
   (*a*) $3\frac{1}{2}$ per cent. of £2 10s. 0d.
   (*b*) 14 per cent. of $3\frac{3}{4}$ tons (answer in cwts.).
6. Find the cost of :
   (*a*) 1 lb. $5\frac{1}{2}$ ozs. of cheese @ 1s. 4d. a lb.
   (*b*) 12 ozs. of fish @ 2s 4d. a lb.
   (*c*) 5 score eggs @ 2s. 3d. a dozen.
   (*d*) 4 yards of cloth @ 5s. 11d. a yard.
   (*e*) 8 cwt. of coal @ £2 7s. 6d. a ton.
   (*f*) 11 gallons of oil @ 1s. $3\frac{1}{2}$d. a gallon.

Comments.

This paper was set to candidates, average age 11 years, who had already been sifted by a Mental Test of their age group. They also took a paper in Arithmetic Problems.

SPECIMEN C. (Time 35 minutes.)

1. Add $\frac{1}{7}$ of 4648 to $\frac{1}{13}$ of 585.
2. Multiply £2 11s. 4½d. by 29.
3. Find the values, in lowest terms, of :

   (*a*) $\frac{5}{6}+\frac{2}{3}$. (*b*) $\frac{1}{3}\times\frac{4}{7}\times\frac{3}{4}$. (*c*) $\frac{3\text{s. } 6\text{d.}}{12\text{s. } 3\text{d.}}$.

4. Find the values of :

   (*a*) $60\times0{\cdot}5\times1{\cdot}2$. (*b*) $3{\cdot}6\div1{\cdot}25$.

   (*c*) $4{\cdot}18\times3{\cdot}75$ (correct to the nearest whole number).
5. Find the values of :

   (*a*) 4½ per cent. of £25.

   (*b*) 18 per cent. of 5½ gallons (answer in pints).
6. Find the costs of :

   (*a*) 6 oz. of tea at 2s. 4d. a lb.

   (*b*) 1 lb. 1½ oz. of cheese at 1s. 4d. a lb.

   (*c*) 5 gallons of oil at 1s. 4d. a gallon.

   (*d*) 2 ft. 3 in. of piping at 4s. 8d. a foot.

   (*e*) 7 drums of paint at 13s. 2d. a drum.

   (*f*) 12 score of eggs at 2s. 1d. a dozen.

   Total cost is not required.

Comments.

This paper was set to candidates, the best third of their age group, who had been selected by Preliminary Tests—(*a*) their school position, and (*b*) an examination. Their average age was 10½ years.

The specimen papers, A, B, and C, are of the same type and require a schedule of marks. For example, C. 1 involves marks for three operations, thus :

$\frac{1}{7}$ of 4648 = 664 (3),

$\frac{1}{13}$ of 585 = 45 (3),

a correct addition (2).

For a slip in either division deduct (2) ; two slips (**0**).

A candidate could make two slips in dividing by 7, divide by 13 correctly and correctly add *his* two quotients and score (**5**).

Question C. 6. Max. (**12**), 2 each, marked right or wrong.

SPECIMEN D. (Time 35 minutes.)

1. Multiply 439 by 7.
2. Divide £7 19s. 11½d. by 11.
3. Take $\frac{1}{9}$ of 585 from twice 67.
4. Reduce £1 18s. 5¼d. to farthings.
5. Reduce 2 tons 11 cwt 1 qr. to lb.
6. How many times 1 ft. 3 in. will measure 7½ yards ?
7. If eggs are 1s. 6d. a dozen, how much are they a hundred ?

Find the values, in lowest terms, of :

8. $1\frac{1}{3} - \frac{5}{6}$. 9. $\frac{1}{4} + (3 \div 1\frac{1}{2})$. 10. $\frac{3\text{s. } 7\text{d.}}{10\text{s. } 9\text{d.}}$.

Find the values, in decimals, of :

11. 3·265 − 0·65.
12. 16 × 1·25 × 0·27.
13. 3·5 + 1·006 − 2·46.
14. What is 16¼ per cent. of 200 tons ?
15. What would 13 oz. of tea cost at 3s. 4d. a lb. ?
16. What would 3¼ yards of cloth cost at 13s. 9d. a yard ?
17. What would 7 lb. of wheat cost at 11s. 8d. a cwt.
18. If John saved 45 per cent. of a sovereign and Mary saved the rest, how many shillings did Mary save ?
19. A man posted some parcels. A third of them weighed 2 lb. each. The rest, 120 in number, weighed 3 lb. each. What was the weight in lb. of all the parcels ?
20. Add together 10s. 4d., 10s. 11d., and 9s. 9d. and divide the total by 3.

Comments.

Paper D was set on a later date at the same type of examination as paper C. The change of form from C to D was suggested by two ideas :

(i) To avoid a constant type of paper in Arithmetic Attainments.

(ii) To reduce the average mark and the number of candidates who scored full marks.

It could not be reasonably anticipated that either C or D would be the invariable form of the paper. The result accorded with plan and, therefore, from the examiner's point of view D is a better type of paper than C ; from the point of view of the schools, perhaps C would be preferred to D. Administratively D has an advantage. The answers are right or wrong, and the marking may be done by a clerk. Some examiners would not care to undertake to mark scripts on papers such as D.

ARITHMETIC PROBLEMS.

SPECIMEN E. (Time 40 minutes.)

1. A line on a drawing half the actual size is $7\frac{1}{2}$ inches in length ; how long would the line have been on a drawing one-third actual size ?

2. A number of children were invited to a Christmas party. Enough oranges were bought for each invited guest to have two. Some children were absent, so that a quarter of those present had a third orange. What fraction of the children who were invited were present ?

3. A full page of a book had 20 lines with 7 words on a line. Chapter III. began on p. 27, which had only 12 lines, and ended on p. 35, which had only 8 lines. The other pages were full. How many words were in Chapter III. ?

4. John and Tom drive a racing motor car in turn for an hour each. John starts ; the first change is made

after 62 miles, the second after 127 miles, the third after 191 miles. Tom finishes, and at the end of 4 hours the car has gone exactly 250 miles. Who drove most miles and by how many ?

Comments.

1. A test (i) in interpretation of " half the actual size," etc., (ii) in the ratio between $\frac{1}{2}$ and $\frac{1}{3}$.
Computation $\frac{2}{3}$ of $7\frac{1}{2} = 5$.

2. See p. 87 *supra*.

3. A test in visualisation of the pagination of a book ; and in the realisation that pp. 27 and 35 equal one full page.
Computation : $8 \times 20 \times 7 = 1120$.

4. A test in visualising the changes. First, the sequence John, Tom, John, Tom. Secondly, the mileages 62, 65, 64, 59. Thirdly, the collection together of the data. John $62 + 64 = 126$,
Tom $65 + 59 = 124$.
Fourthly, the answer John by 2 miles.
The scripts were extraordinarily interesting. There were answers inconsistent in themselves—" Tom by 2 miles."
There were some attempts which broke down on the sequence only, others which broke down on the mileages only. This question worked out very successfully.

SPECIMEN F. (Time 40 minutes.)

1. A bus conductor worked for four hours. The second hour he took twice as much in fares as the first hour ; the third hour three times as much as the second hour ; and the last hour four times as much as the third hour. He took in all £9 18s. 0d. ; how much did he take the second hour ?

2. A year ago a man bought some land for £100 and, at a cost of £1000, built a house on it. He spent also £60 for making the garden and £120 for decorations. (*a*) What was the total cost ?

This year land is half as dear again; buildings cost 25 per cent. more; and decorations cost 10 per cent. more. The garden can be made for £50. (*b*) What would be the cost this year?

3. A man has 2 cwt. of tea which is to be weighed into 1 lb. packets. He always put at least 1 lb. in each packet, but may put as much as 1 lb. and $\frac{1}{4}$ oz., because his scales are not quite exact. What is the smallest number of packets he can make and yet have no tea left?

4. A pole is painted black for equal distances from the top and from the bottom. The rest is painted with an equal number of red and white bands. A red band is twice as wide as a white band. If one black piece is $\frac{1}{8}$ of the pole, what fraction of the pole is white?

5. 27 white wooden blocks each 1″ by 1″ by 1″ are built to make a solid 3″ by 3″ by 3″.

The top and sides of the solid are then painted red. The blocks are then separated and sorted into piles:

(*a*) those with three red faces,
(*b*) those with two red faces,
(*c*) those with one red face,
(*d*) those which are still all white.

How many blocks are there in each pile?

Comments.

1. A test in English, and in selection of data.
Computation: $1+2+6+24 = 33$,
$\frac{2}{3}\frac{2}{3}$ of £9 18s. 0d. = 12 shillings.

2. A test in continuity of effort, and in sorting out data.
Computation:

(*a*)	(*b*)
£100	£150
1000	1250
60	50
120	132
1280	1582

3. A rather difficult test in visualising a grocer weighing tea. The word "smallest" implies the "largest" possible number of packets each 1 lb. plus $\frac{1}{4}$ oz., and, therefore, the notion: $64\frac{1}{4}$ packets contain 65 lb. of tea.

4. Visualisation of the painted pole should lead to a sketch and an immediate solution: $\frac{1}{3}$ of $\frac{3}{4} = \frac{1}{4}$.

5. Visualisation of a solid to permit thinking of areas in reference to familiar wooden blocks.

(*a*) Must be corners at the top. (4)

(*b*) Must be top and one side, or two sides, or bottom corners. (12)

(*c*) Must be in the middle of the top or a side and in middle of bottom edges. (9)

(*d*) Those left out of 27, hidden. (2)

SPECIMEN G. (Time 40 minutes.)

1. A man earned £5 0s. 9d. in a week of $46\frac{1}{2}$ working hours. On each working day from Monday to Friday he worked the same number of hours. On Saturdays he worked 4 hours. For how long did he work on Wednesdays ?

2. Nine girls and a forewoman work in a factory ; they have 45 minutes allowed for dinner. Three girls go for dinner at 12 noon ; three others at 12.30 p.m. and three others at 1 p.m. The forewoman goes at 1.15 p.m. How many are there at work in the factory at 1.10 p.m. ?

3. One side of an oblong piece of wood is painted with three black and two red stripes. The black stripes are the same width and each red stripe is twice the width of a black stripe. If the total black area is 12 square feet, what is the area of one red stripe ?

4. Tom had a bag of marbles, of which $\frac{1}{3}$ were glass and the rest stone. Jack has as many marbles as Tom but $\frac{1}{3}$ of his were stone and the rest glass.

They mixed the marbles and then each took half of them. For every 3 glass marbles which Tom now had

he had 7 stone marbles. What fraction of Jack's marbles were glass ?

5. A car which travelled 21 miles per hour was to leave "The Laurels" in time to meet Mr. White at the station and bring him home by 6 p.m. The car was 5 minutes late in starting and brought Mr. White home at 5.45 p.m. Mr. White had caught an earlier train and had walked homewards for an hour before he met the car. At what speed did Mr. White walk ?

Comments.

1. A "slice of life" to be thought about.

Computation : $\frac{1}{7}$ of $(46\frac{1}{2} - 4) = 8\frac{1}{2}$.

2. Another "slice of life" to be sorted out piecemeal. A test in English on the word "forewoman."

Possible answers : 0, 3, 4, 6, 7, 9, 10.

No computation.

3. Visualisation to be helped by a diagram or sketch.

The sketch is not too easy since only *area* is given.

Computation : $\frac{1}{4}$ of 12 sq. ft. $= 3\frac{1}{4}$ sq. ft.

4. Visualisation by stages :

(i) Altogether as many glass as stone.

(ii) Tom : 3 glass 7 stone,
Jack : 7 glass 3 stone.

5. Yet another "slice of life"; involving the notion of the double journey "there and back" of the car.

Car travelled 20 mins. less than usual.
,, ,, 7 miles ,, ,, ,,
Mr. White ,, $3\frac{1}{2}$,, ,, ,, ,, *in the car:*
Answer : $3\frac{1}{2}$ miles per hour.

Papers F and G contained five, but E only four problems. All three were set at the examination in different years for the same authority's scholarships, and were tackled by children of $10\frac{1}{2}$ years.

Paper E and similar papers of four questions were not testing adequately the most capable children ; relatively too many were scoring full marks.

Papers F and G were then tried with five questions, in the same time as before, and even these fail to test the most capable children to the maximum ; still too many worked these five problems correctly in 40 minutes. Of course, there were fewer with full marks than in E, but the numbers who scored the possible maxima on F and G were surprising.

SPECIMEN H. (Time 40 minutes.)

1.

County	Area in square miles	Number of sheep
A	1334	444,065
B	875	437,780

Express the number of sheep per square mile in county *B* as a percentage (%) of the number of sheep per square mile in county *A*.

2.

	Column 1	Column 2
	s. *d.*	*s.* *d.*
A	37 6	50 0
B	— —	— —
C	18 6	— —
Total	66 0	— —

Column 2 is to show the same proportional increases over Column 1 throughout.

Fill in the blanks in the table and copy them in your answer book.

3. The speedometer of a motor car is slightly inaccurate. From mile-post 31 to mile-post 46 it registers 16 miles ; how far has the car really gone when the speedometer has registered 88 miles ?

4. A man's rate of wages is 1s. 10d. an hour ; overtime is paid for at $1\frac{1}{4}$ times the usual rats. He works 44 hours

a week and overtime as follows: Monday $2\frac{3}{4}$ hrs.; Tuesday $1\frac{1}{2}$ hrs.; Friday $1\frac{3}{4}$ hrs.; Saturday 2 hrs. How much does he earn in the week?

5. A piece of white cardboard is 8″ by 6″. Strips of black paper 1″ wide are gummed round the cardboard to make a black border an inch wide.

(*a*) What length of black paper is required?

(*b*) What fraction of the original white surface is changed to black?

Comments.

Paper H was set to children aged 13 + years.

1. A test in the recognition of ratios. If the candidates had been in the habit of making rough estimates of the answers to calculations they might have noticed that *A* came to 333, and *B* to 500 sufficiently closely to save computation, and so have finally written 500 on 333 is 50 per cent. increase.

2. An extraordinarily good selective test. In most scripts there was no evidence of any connection in the minds of the candidates between 37s. 6d. and 50s., and the words "same proportional increases."

Computation: 50s. is a third increase on 37s. 6d.

B 10s. and 13s. 4d., *C* 24s. 8d. Total 88s.

3. A simplified "slice of life."

Computation: $\frac{1}{16}$ of $88 = 5\frac{1}{2}$. Answer $82\frac{1}{2}$.

4. Checking a "wages" envelope.

Computation: Addition of overtime, 8 hrs.,
54 times 1s. 10d. = £4 19s. 0d.

5. A test, primarily, of ability to present two consistent answers.

ENGLISH.

Specimen Essay Subjects.

In Essay papers a choice of from two to four subjects is usually given. The following subjects have been set.

Candidates aged $10\frac{1}{2}$ years. (Time 40 minutes.)

1. Write about Ships and their uses.

2. Make up a story about the following people : Jack and his little sister Dorothy, with Jim their dog ; Ben Brown, an old man of eighty ; and Betty.

3. Once upon a time a king sent his three children away to spend a year in seeing the world. The eldest had cunning, the youngest great strength, and the other was an ordinary mortal.

Describe what happened to *one* of the three.

4. Different kinds of roads and where they lead to.

Comments.

It is not desirable that the Essay should always be written without some assistance. For example, Whaling as a subject would possibly not be sufficiently familiar to the candidates to give them enough ideas for a prose exercise of adequate length. In such cases assistance may be given as in the following :

Rubric.

You will be given a story told in verse. Read it once to find what it is about. Read it again to find the sequence of events. Read it a third time to find all the points of the story, and to find a title.

At the end of 15 minutes the story will be collected and you will be told what to do.

The Story.

In seventeen hundred and ninety-four,
On March the twentieth day,
We hoist our colours to the mast,
And for Greenland bore away, brave boys!
And for Greenland bore away.

We were twelve gallant men aboard,
And to the North did steer:
Old England left we in our wake—
We sailors knew no fear, brave boys!
We sailors knew no fear.

Our boatswain to the mast-head went,
Wi' a spy glass in his hand;
He cries, 'A whale! a whale doth blow,
She blows at every span, brave boys!
She blows at every span.'

Our Captain on the master deck
(A very good man was he),
'Overhaul! overhaul! let the boat tackle fall,
And launch your boat to sea, brave boys!
And launch your boat to sea.'

Our boat being launch'd, and all hands in,
The whale was full in view;
Resolved was then each seaman bold
To steer where the whale-fish blew, brave boys!
To steer where the whale-fish blew.

The whale was struck, and the line paid out,
She gave a flash with her tail;
The boat capsized and we lost four men,
And we never caught that whale, brave boys!
And we never caught that whale.

Bad news we to the Captain brought,
The loss of four men true;
A sorrowful man was our Captain then,
And the colours down he drew, brave boys!
And the colours down he drew.

'The losing of this whale,' said he,
'Doth grieve my heart full sore;
But the losing of four gallant men
Doth hurt me ten times more, brave boys!
Doth hurt me ten times more.

'The winter star doth now appear,
So, boys, the anchor weigh;
'Tis time to leave this cold country,
And for England bear away, brave boys!
And for England bear away.

'For Greenland is a barren place,
A land where grows no green,
But ice and snow, and the whale-fish blow,
And the daylight's seldom seen, brave boys!
And the daylight's seldom seen!

The oral instructions. Write in your own words the story you have read.

ENGLISH QUESTION PAPERS.

SPECIMEN J. (Time 30 minutes.)

1. Re-write this paragraph, using the exact words spoken by the speakers:

A man and a boy were walking along a lane. The man asked the boy to run to the field and call the cows, but the boy said that he had hurt his leg and could not run. Thereupon the man said that he would call the cows himself.

2. Read the following verse of poetry very carefully:

The boy stood on the burning deck,
Whence all but he had fled,
The flames that lit the battle's wreck,
Shone round him o'er the dead.

Then answer the following questions :

(i) How many men and boys were left alive on deck ?

(ii) What could be seen by the light of the burning ship ?

(iii) What is meant by ' wreck ' ?

3. Each of the following sentences contains the word *etc.* Make for each sentence a list of four words which might be used instead of *etc.*

(*a*) Carrots, cabbages, etc., were exposed for sale on the stall.

(*b*) Toys, books, etc., were littered about the classroom.

(*c*) A farmer owns many cows, pigs, etc.

(*d*) A tram driver requires patience, etc.

(*e*) Traffic includes motor cars, carts, etc.

4. Complete the following sentences in a suitable manner :

(1) Her eyes shone like —— .

(2) Robin Hood was as brave as —— .

(3) The sky grew black as —— .

(4) On the grass the dewdrops glittered more brightly than —— .

(5) The dog stood as still as if —— .

Comments.

This paper was set in a preliminary examination for children of 10 years. The candidates competed solely against their own class-mates and the scripts were marked by their own Head Teacher.

SPECIMEN K. (Time 35 minutes.)

1. Read the following passage :

I've heard about a pleasant land,
Where omelettes grow on trees,
And roasted pigs run, crying out,
' Come eat me, if you please.'

My appetite is rather keen,
But how shall I get there ?
' Straight down the Crooked Lane,
And all round the Square.'

Answer the following questions :

(*a*) Invent a name for the pleasant land.

(*b*) Point out three absurdities in the passage.

(*c*) Quote the reason why I want to reach the pleasant land.

(*d*) Express in your own words the directions I am to follow to reach the pleasant land.

2. Read the following passage :

The gigantic steamer entered the dock and proceeded to discharge her cargo. Meanwhile the captain went ashore and reported his safe arrival.

Write down, one on a line, the words underlined and opposite each of these words write a word or phrase which has a similar meaning. For example :

gigantic, huge.

3. Read the following phrases :

(*a*) at once, (*b*) solitary,

(*c*) not an arm's length off, (*d*) perfect,

(*e*) in need of funds, (*f*) in great glee,

(*g*) in a clumsy manner.

For each there is a phrase beginning with one of the words *with, within,* or *without,* which has almost exactly the same meaning.

Choose five of the phrases and write each in your answer book with its equivalent, in this manner :

unafraid,
without fear.

4. Read these sentences :

(*a*) Grasp the —— firmly.
—— it gently.

The word omitted is the same word—handle.

Ten pairs of sentences follow, write in your answer book the missing word, thus, (*a*) handle.

(*b*) My —— is John.
A dog with a bad —— .

(*c*) We got in after a long —— .
Time and tide —— for no man.

(*d*) Whose —— is it ?
I can't —— ends meet.

(*e*) Watch him —— rapidly in the air.
This is only a gentle —— .

(*f*) He was in fine —— .
—— your nest for the future.

(*g*) This boot hurts my —— .
I saw the yacht —— over and capsize.

(*h*) The —— scored five only.
I'm tired, I must —— .

(*i*) The present was a tiny gold —— watch.
He fell from his horse and broke his ——

(*k*) The singer had a —— reception.
Wrap up, or you'll catch —— .

(*l*) Don't give yourself —— .
Try these —— over at the piano.

Comments.

This paper was set at an open scholarship examination for candidates aged 10½ years.

Question 4 was a severe test in concentrated effort and was a good selective question.

SPECIMEN L. (Time 35 minutes.)

1. Read the following :

And thereupon we all entered the cave. It was a large, airy place, with a little spring and a pool of clear water, overhung with ferns. The floor was sand. Before a big fire lay Captain Smollett ; and, in a far corner, only duskily flickered over by the blaze, I beheld great heaps

of coin and quadrilaterals built of bars of gold. That was Flint's treasure that we had come so far to seek, and that had cost already the lives of seventeen men from the *Hispaniola*. How many it had cost in amassing, what blood and sorrow, what good ships scuttled on the deep, what brave men walking the plank blindfold, what shot of cannon, what shame and lies and cruelty, perhaps no man alive could tell.

Answer these questions :

(*a*) What was Flint ?

(*b*) What was the *Hispaniola* ?

(*c*) State in your own words what 'I beheld.'

(*d*) Express in your own words the meaning of :

(i) duskily flickered over by the blaze ;
(ii) scuttled on the deep ;
(iii) walking the plank.

(*e*) How many lives had the treasure cost ?

2. Here are four words :

wound, race, pitch, play.

Choose *two* of them and write *six* sentences (three for each word) to show three different ways in which the chosen word may be used.

3. Write the following in your own words :

Gibraltar.

Seven weeks of sea and twice seven days of storm
Upon the huge Atlantic, and once more
We ride into still water and the calm
Of a sweet evening, screen'd by either shore
Of Spain and Barbary. Our toils are o'er,
Our exile is accomplished. Once again
We look on Europe.

Read the following phrases. For each phrase there is another phrase which means almost exactly the same and contains one of the words—*after, before, behind.*

Choose four of the phrases and write each with your answer in this manner :

ere his task was completed
before his work was done

(*a*) when night had fallen,
(*b*) at the tail of the cart,
(*c*) previous to his coming,
(*d*) face to face with the judge,
(*e*) last of all,
(*f*) so soon as he was acquitted.

Comments.

Candidates who were successful with Specimen J were given Specimen L in the subsequent open scholarship examination when they were barely six months older.

SPECIMEN M. (Time 45 minutes.)

1. Read the following :

All day long the traffic goes
In Lady Street by dingy rows
Of sloven houses, tattered shops—
Fried fish, old clothes and fortune tellers—
Tall trams on silver-shining rails,
With grinding wheels and swaying tops,
And lorries with their corded bales,
And screeching cars.

Answer the following :

(*a*) *Quote* the words which describe the buildings.
(*b*) Express in your own words ' silver-shining ' and ' corded bales.'
(*c*) Where are the old clothes ?
(*d*) Why do the tops sway ?
(*e*) *Quote* the words which give you the idea of noise.

2. Here are three words :

hand, jar, port.

Write *nine* sentences, three for each word, to show three different ways in which each word may be used.

3. Each of the following indicates movement in a certain direction:

(*a*) Bear to the left.
(*b*) The swallows wing their way southwards.
(*c*) Alight from the train.
(*d*) They walked along the deck towards the stern of the ship.
(*e*) The down train was due in three minutes.

Write on your paper against the letters *a*, *b*, etc. (one on a line), the word or words to use in these sentences to show that the movement is in the opposite direction.

4. Express in your own words :

(*a*) Fit as a fiddle. (*b*) Safe bind, safe find.
(*c*) A stitch in time saves nine.

5. Here are eleven numbered phrases :

(1) Without a sound. (2) At once.
(3) As one man. (4) Without fear.
(5) On the very top. (6) Like a thief.
(7) With every care for others.
(8) A bit here and a bit there.
(9) Without mercy. (10) Draw near to.
(11) Adjudged not guilty.

In place of each of the above phrases a single word could be used which has the same meaning, *e.g.* (1) soundlessly and without a sound ; but noiselessly is a better example than soundlessly because soundlessly repeats the given word ' sound.'

Write the numbers 1 to 11 on your paper, one on a line, and write against each number the word to use instead of the phrase,

e.g. (1) noiselessly.

Comments.

This paper was set to candidates aged 11 years, who had also Intelligent Test papers in Language and Number.

Question 5 is a question of Intelligence Test type in that the answers are single words instead of phrases or sentences. It would not be admissible as a Mental Test question because it is not possible to assert beforehand in each case 'this word is right, all other words are wrong.' The examiners make a list during their preliminary survey of all the words submitted by the candidates, at the conference of examiners they decide (with special reference to the frequency with which each word appears on the scripts) which words shall be accepted; in some cases three or more words score equally well. The scoring depends largely upon how the question has appealed to the candidates, and the schedule of marking would probably differ for two different examinations.

The other questions are similarly marked objectively.

SPECIMEN N. (Time 35 minutes.)

1. Read this passage:

'My Lords of Sussex and Leicester, we require your presence at the Privy-Council to be presently held where matters of importance are to be debated. We will then take the water for our divertisement, and you, my lords, will attend us.—And that reminds us of a circumstance—Do you, Sir Squire of the Soiled Cassock (distinguishing Raleigh by a smile), fail not to observe that you are to attend us on our progress. You shall be supplied with suitable means to reform your wardrobe.'

Answer these questions:

(*a*) What is a cassock, and who owned it?

(*b*) What did Queen Elizabeth propose to do after the meeting of the Privy-Council, and why?

(*c*) *Quote* the words which state the purpose of the meeting of the privy-council.

(*d*) What command did the Queen lay upon Raleigh?

2. Here are eleven words, each of which fits into one of the numbered spaces in the passage below :

Words—last, endure, walk, pain, grin, nothing, jeers, something, display, stocking, precipitation.

Passage—I am by nature extremely susceptible of street affronts : the (1) and taunts of the populace ; the low-bred triumph they (2) over the casual trip, or splashed (3) , of a gentleman. Yet can I (4) the jocularity of a young sweep with (5) more than forgiveness. In the (6) winter but one, pacing along Cheapside with my accustomed (7) when I (8) westward, a treacherous slide brought me upon my back in an instant. I scrambled up with (9) and shame enough—yet outwardly trying to face it down, as if (10) had happened—when the roguish (11) of one of these young wits encountered me.

(Make a *list* of the numbers from 1 to 11 in your answer book and write against each number the word which fits.)

3. Here are two columns of phrases. In the second column there is one phrase which means a greater quantity or degree of something suggested by a phrase in the first column. For example (4) means faster movement than (*g*) :

(*a*) Minor ailment.	(1) Much ado about nothing.
(*b*) Passing fancy.	(2) First rate.
(*c*) Scarcely polite.	(3) Absorbing attachment.
(*d*) A little fuss over a trifle.	(4) With lightning speed.
(*e*) Fairly well.	(5) Whole hearted merriment.
(*f*) A laugh or two.	(6) Serious illness.
(*g*) At a moderate pace.	(7) Thorough gentleman.
(*h*) Chance acquaintance.	(8) Windy oration.
(*j*) Milk and water criticisms.	(9) Lifelong comrade.

Find the phrases which make a similar pair with

(*a*), (*b*), (*d*), (*f*), (*h*),

and write the pairs in your answer book, one pair on a line in this way :

(*g*) and (4).

4. Explain briefly in your own words the following :
   - (*a*) A miss is as good as a mile.
   - (*b*) Without fear or favour.
   - (*c*) In the twinkling of an eye.
   - (*d*) In a slough of despond.
   - (*e*) Lackadaisical.

5. Here are ten numbered phrases. Read them :

(1) In the near future.
(2) At all costs.
(3) Kith and kin.
(4) As merry as a marriage bell.
(5) Within an ace of disaster.
(6) To the top of his bent.
(7) A queer specimen.
(8) By slow degrees.
(9) In the grip of fear.
(10) At a moment's notice.

Write the numbers 1 to 10, one on a line of your answer book. Write against each number a word or phrase which has the same meaning as the corresponding phrase in the question paper. *The shorter the phrase the better.* For example—for the phrase (1) *In the near future, soon* or *to-morrow* would be a better answer than *in a day or two.*

Comments.

This paper was set to candidates aged 13 years, who were also submitted to a paper in History.

Question 2 is an Intelligence Type question which could also be used in a Mental Test. The marking is objective in this examination, while in a Mental Test the marking would take the form of a score in comparison with a normal score for the age group previously and independently determined.

SPECIMEN O. (Time 35 minutes.)

1. Read the following passage :

Beside yon straggling fence that skirts the way,
With blossom'd furze unprofitably gay,
There, in his noisy mansion, skill'd to rule,
The village master taught his little school :
A man severe he was, and stern to view,
I knew him well, and every truant knew :
Well had the boding tremblers learn'd to trace
The day's disasters in his morning face ;
Full well they laugh'd with counterfeited glee
At all his jokes, for many a joke had he.

Answer the following :

(*a*) Express in your own words the meaning of ' skirts the way ' ; ' noisy mansion ' ; ' with counterfeited glee.'

(*b*) Express the meaning of the seventh and eighth lines in your own words.

(*c*) Quote the *four* words which describe the schoolmaster.

(*d*) What did every truant know ?

2. Here are five phrases :

(*a*) An unpleasant necessity.
(*b*) A baffling problem.
(*c*) A meritorious win.
(*d*) A deathlike silence.
(*e*) A lame excuse.

Write *five* sentences in each of which one of these phrases is suitably used.

3. Here are eleven words, each of which fits into one of the numbered spaces in the passage below :

Words—silver, breeds, fish, gravel, ground, feasts, body, excellent, wholesome, commended, feeding.

Passage—The gudgeon is reputed a (1) of excellent taste and to be very (2) . He is of a fine shape, of a

(3) colour, and beautified with black spots both on his (4) and tail. He (5) two or three times in the year, and always in summer. He is (6) for a fish of (7) nourishment. The Germans call him Groundling, by reason of his (8) on the (9) ; and he there (10) himself, in swift streams and on the (11) .

Make a list of the numbers 1 to 11 in your answer book and write against each number the word which fits.

4. Here are five words :

(*a*) mere, (*b*) passing, (*c*) cherished,
(*d*) embroidered, (*e*) brilliant.

Write in your answer book the letters (*a*) to (*e*). Against each letter write *two* phrases each of which contains the word printed above ; for example, against (*a*) write *two* phrases in which the word 'mere' has different meanings.

5. Here are ten incomplete phrases. Write them in your answer book, and complete each by the addition of a word or words so as to make sense of each :

(1) As quick as ——. (2) As greedy as ——.
(3) As silent as ——. (4) As tender as ——.
(5) As soft as ——. (6) As rich as ——.
(7) As —— as a tortoise. (8) As —— as coal.
(9) As —— as ice. (10) As —— as iron.

Comment.

Another paper for the same range of candidates as Specimen N.

HISTORY.

Specimen P. (Time 1 hour.)

1. Write for about twenty minutes upon the life and times of *one* of the following :

Canute, Henry VIII., Sir Robert Peel.

2. Select *one* of the following events :

Treaty of Wedmore.
The Battle of Barnet.
The Opening of the Suez Canal.

Write about it as if you were present and so describe the circumstances which led up to the event and what actually happened.

3. Answer *four* of the following :

(*a*) Name three English leaders who led English (or British) armies to victory on the Continent.

(*b*) Name three English leaders who led English (or British) armies to victory in parts of the world outside Europe.

(*c*) Name three English leaders who won great naval victories.

(*d*) Name three English sovereigns who were the first kings of their line.

(*e*) Name three English sovereigns who ruled during the eighteenth century.

(*f*) Name three English sovereigns who ruled over portions of Europe.

4. Choose *six* of the following eight events and arrange them in historical order, beginning with the most recent :

The Battle of Trafalgar.
Captain Cook sailed to New Zealand.
Jack Cade's rebellion.
The English victory at Crecy.
' Remove that bauble,' ordered Cromwell.
The first train ran between Liverpool and Manchester.
The death of Becket.
Richard I. led the Third Crusade.

Comments.

This paper was set to scholarship candidates aged about 13 years.

Questions 1 and 2 are Essay Type questions with such options as are desirable on a very general syllabus.

Questions 3 and 4 are Intelligence Type questions intended to cover the whole range of the subject in relation to its more salient features.

GEOGRAPHY.

SPECIMEN Q. (Time 1½ hours.)

1. On the blank map of the British Isles write your examination number in the top right-hand corner, and

(*a*) Name, without boundaries, East Anglia, Fife, Glamorgan.

(*b*) Mark and name: the Cumbrian Mountains, the South Downs, the Grampians.

(*c*) Insert and name: the Severn, the Tay, the Mersey.

(*d*) Insert and name: Belfast, Cork, Chester, Carlisle, Dundee, Edinburgh, Bristol, Hull.

2. Write about *either* the West Riding *or* the Black Country under the following heads:

(*a*) Where it is.

(*b*) What big towns are there.

(*c*) The chief work of the people in factory and field.

(*d*) The chief railways of the district and the goods they carry.

3. Write short answers to *five* of the following:

(*a*) What sort of vegetation occurs in the Amazon basin?

(*b*) What wild land animals are hunted by the Eskimo?

(*c*) What mineral mined in England will dissolve in water and where is it mined?

(*d*) Why do the glaciers in Greenland reach the sea-shore while the glaciers of the Himalayas remain high up in the mountain valleys?

(*e*) Which is usually the warmest during January: Glasgow, Newcastle or Dover ?

(*f*) Which town is highest above sea-level: York, Hull or Stoke-on-Trent ?

(*g*) Why are there few people in Central Australia ?

(*h*) Which crop is grown extensively in the Punjab: tea, wheat or rice ?

4. Answer *four* of the following:

(*a*) Name three seaports in America on the Pacific Ocean.

(*b*) Name three large towns in the same country as Rome.

(*c*) Name three islands in the Mediterranean Sea.

(*d*) Name three British portions of Africa.

(*e*) Name three other islands in the same group as Jamaica.

(*f*) Name three countries each of which contains a portion of the Alps.

5. Answer *four* of the following:

(*a*) Name an English district where some people mine coal, others work in iron and others make earthenware.

(*b*) Name an Irish district where some people build ships, others grow oats and flax and others make sheets, table cloths, etc.

(*c*) Name a Canadian district where some people grow fruit, others cut down trees, and others catch and tin fish.

(*d*) Name an island in the Mediterranean which contains a volcano, is visited by earthquakes, and sends lemons to England.

(*e*) Name a British district in the Far East where some people mine tin, others grow rubber trees, and most eat rice.

(*f*) Name an island in the British Empire where some people grow tea, others grow coconut palms, and only a few people are white.

Comments.

This paper was set to candidates aged 13 years.

Question 1 is a matter of fact question, and the examiners decide objectively upon the limits of permissible location of places, etc.

Question 2 is an Essay Type question with a limitation of the main features of the answer.

Questions 3, 4, and 5 are Intelligence Type questions, some of which would serve as Mental Test questions.

INDEX

PRINTED IN GREAT BRITAIN BY ROBERT MACLEHOSE AND CO. LTD
THE UNIVERSITY PRESS, GLASGOW.